LA-UR--97-2534

MASTER

Celebrating the Neutrino

Los Alamos
NATIONAL LABORATORY

S Los Alamos Science

DISCLAIMER

DISCLAIMER

Portions of this document may be illegible electronic image products. Images are produced from the best available original document.

Celebrating

*Receiving the Nobel
Prize for the detection of
the neutrino is a great honor
not only for Fred Reines
personally but also for
Los Alamos National
Laboratory—this is the first
time Laboratory-sponsored
work has received such
recognition. The Laboratory
is extremely proud of Reines
for bringing that honor to it.*

In 1995, Fred Reines was awarded the Nobel Prize in physics for the detection of the neutrino, perhaps the most intriguing and certainly the most elusive of nature's elementary particles. Wolfgang Pauli had "invented" the neutrino in December 1930, when the only known particles were the proton, the electron, and the photon, and the contents of the nucleus were still a mystery. The new particle was to carry away the missing energy in beta decay, the nuclear decay process in which a nucleus of one element expels an electron and changes to the nucleus of the next higher element in the periodic table. The invisible particle accompanying the electron had to have the same spin as the electron, but little or no mass, and it had to be electrically neutral. Most importantly, it had to interact with matter very very weakly to explain why it was never observed in the experiments of the day.

Pauli viewed this hypothesis as a desperate attempt at saving the time-honored law of energy conservation. But a few years later, Enrico Fermi used the concept of the neutrino to develop his very successful theory of beta decay. By association, the neutrino became real in the minds of most physicists even though this particle was believed to be impossible to detect. Then, in the 1950s, Reines, Clyde Cowan, Jr., and their Los Alamos team finally showed that the neutrino could be observed away from its point of origin. Their ground-breaking experiments changed the role that the neutrino was to play in physics. It would no longer be merely the invisible byproduct of beta decay, a passive partner to the electron to be accounted for and then ignored, but a useful tool for uncovering the basic structure of matter. The neutrino's fascinating "career" in physics and long tenure in the experimental programs at Los Alamos provides the focus of the present volume.

For all its staying power in physics, the neutrino is not a household word. The reason is simple: The neutrino interacts with matter only through the weak force, the least known of nature's four fundamental forces. Compared with the other three, the weak force seems like nature's afterthought. Gravity holds us to Earth and keeps the planets in their orbits. The electromagnetic force holds atoms together and governs the chemistry of the elements. And the strong force binds protons and neutrons into nuclei and governs the nuclear processes of fission and fusion. But what role does the weak force play?

All of us learn that fusion produces the energy that powers the stars and turns hydrogen into helium and helium into heavier elements. But not all of us are aware that the weak force is at work in these processes. Beginning with the fusion of one proton with another to make deuterium and up through the periodic table—from hydrogen to helium, from helium to carbon, and so on—the making of the elements includes not only the merging of nuclei through the strong force but also the slower transmutation of one element into another through the weak force. In each case, the weak force causes a neutron to transform into a proton or a proton into a neutron. Simultaneously, it keeps electric charge constant by creating or destroying an electron (or its antiparticle, the positron). And it keeps something called weak charge constant by creating or destroying the neutrino (or the antineutrino), the electron's weak partner.

Fred Reines

the Neutrino

The weak force began to shape the universe during the first hundred seconds following the Big Bang. Neutrons and protons were in thermal equilibrium, and they interacted through the weak force (as well as the strong) to create the primordial abundances of helium and other light elements. Today the weak force continues to shape our world, causing the Sun to shine for us by day and the stars to twinkle in the heavens by night. Whether in the interior of stars or in the expanding envelopes of supernovae, the weak force is at work making all the elements we know on Earth. It is the force of transmutation. Without it, the elements, and therefore we, would never have come to be.

And so, the neutrino's importance becomes apparent: Wherever the weak force plays a role, we usually find the neutrino either as a relic of the interaction or as a major participant. We are permeated by a thermal background of neutrinos, which, like the background of cosmic microwave radiation, is a leftover from the Big Bang. These neutrinos decoupled from the rest of matter about one second after the Big Bang and continued to expand and cool on their own. There are about three hundred of them in every cubic centimeter of our universe. Because of Earth's proximity to the Sun, we are also drenched by a giant flux of neutrinos from the fusion reactions in the Sun's core: roughly 600 billion solar neutrinos go through our fingertips every second. But these elusive neutrinos have such a small chance of interacting with matter that their presence goes mostly unnoticed. They pass silently through everything they meet as they begin an almost endless journey through the solar system, past our galaxy, and out to the far reaches of the universe. Because neutrinos interact so little, however, they preserve a record of the conditions present at the place of their creation—whether at the center of the Sun or at the center of a dying star on its way to becoming a supernova.

Many physicists have followed in the footsteps of Reines and Cowan, building oversized detectors to trap the neutrinos and decode their messages. The neutrino's intrinsic properties were measured, and it was discovered that three distinct types, or flavors, of neutrino exist. High-energy neutrinos and antineutrinos were used to probe the substructure of neutrons and protons and to confirm (in conjunction with electron scattering experiments) that quarks have fractional charge and that triplets of quarks make up each neutron and proton. Neutrino interactions with matter were also used to demonstrate that the weak force sometimes acts like the electromagnetic force, causing particles to scatter without transmuting their identities. Time after time, the neutrino has been used to expand our understanding of the subatomic world.

Today, the neutrino is at center stage in particle physics as scientists approach a definitive answer to a most basic question: Do neutrinos have mass? The positive results from the Los Alamos neutrino oscillation experiment (called LSND for the name of the neutrino detector) suggest that the answer is "yes." Neutrino oscillations are a completely nonintuitive phenomenon in which the different neutrino flavors shift their identities and transform into one another. The shift is not caused by the weak force but by an interference phenomenon that can occur only if at least one neutrino flavor has mass. Thus LSND's positive results are of huge interest to physicists all over the globe. Confirmation of these results would finally prove that neutrinos have mass. It would also prove that neutrinos

Clyde Cowan, Jr., and Fred Reines led the team of talented scientists and technicians at Los Alamos that made the first definitive measurement of an event induced by a free neutrino, thus proving that this theoretical construct did indeed exist.
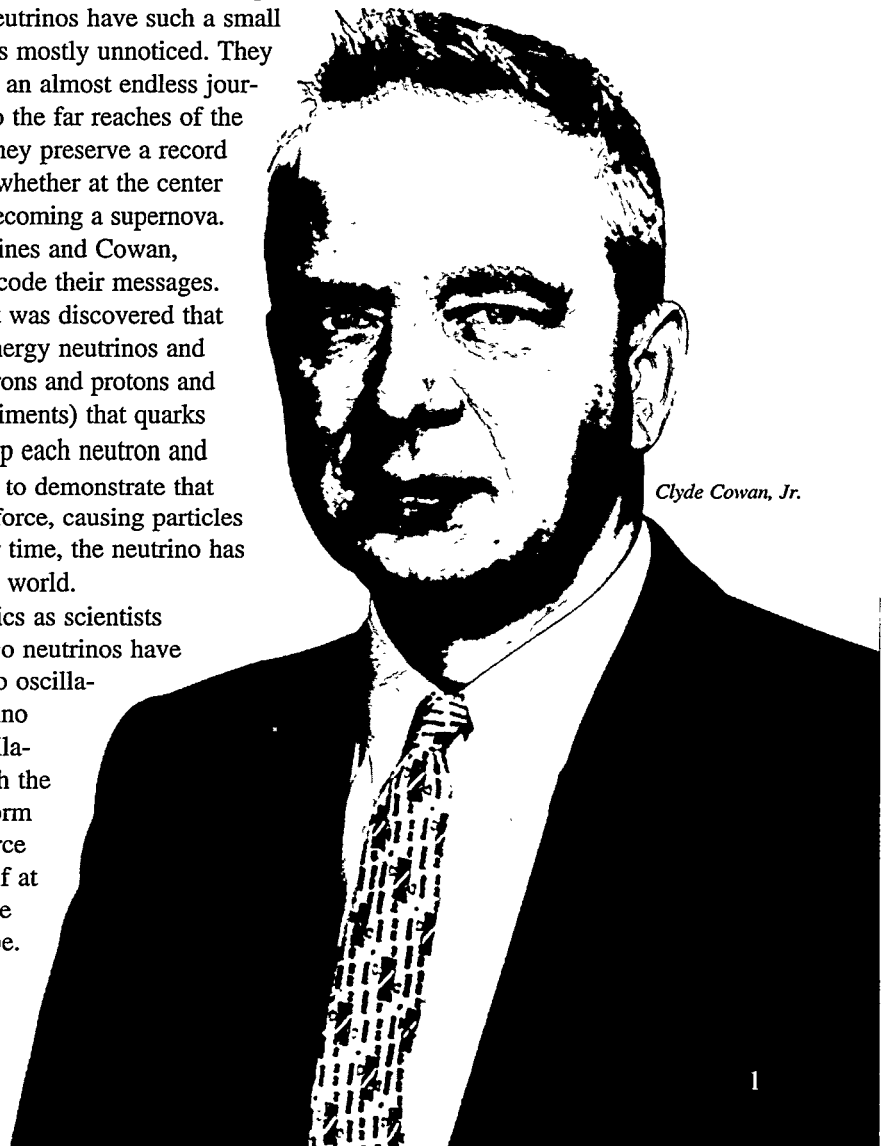
Clyde Cowan, Jr.
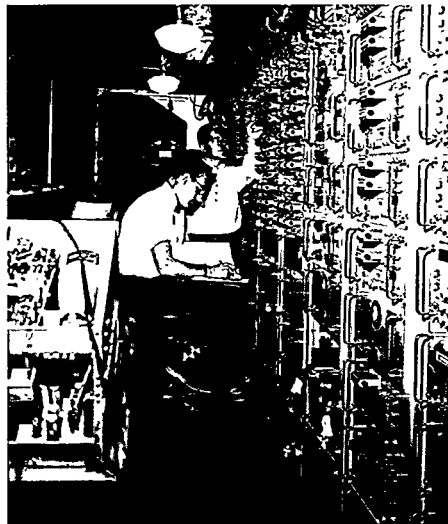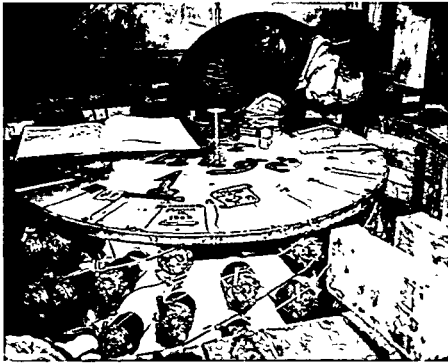
mix among their different flavors in the same way that quarks mix among their different flavors.

The added similarity between the quarks (strongly interacting particles) and the neutrinos (weakly interacting particles) would increase the hope that the world of elementary particles can be described by a Grand Unified Theory. This popular extension to the Standard Model of particle physics describes three of nature's forces as different aspects of a single force, and the neutrinos and all the elementary particles as different excitations of a single quantum field. The existence of neutrino oscillations would also suggest that this phenomenon is the explanation for the solar-neutrino puzzle, the measured shortfall in the flux of neutrinos coming from our Sun.

On a more local note, confirmation of the LSND results would mean that Los Alamos scientists have once again performed a historic experiment in the field of neutrino physics. Such an achievement would be a natural outcome of the ongoing commitment at Los Alamos to this particular field. Ever since the Reines and Cowan experiments, the Laboratory's scientists have exploited many resources to advance neutrino studies. The Los Alamos Neutron Science Center (formerly called LAMPF) became a tool to study the differences between muon and electron neutrinos. The Laboratory's expertise in isotopes and very low level radioactivity was applied to search for neutrinoless double beta decay, a process that would prove that the neutrino is a Majorana particle (a particle that is its own antiparticle). The Laboratory's nearly unique capacity for handling tritium was applied in a precedent-setting search to measure the mass of the electron neutrino directly. Later, Los Alamos scientists went beyond the Laboratory to join American and Russian colleagues in a solar-neutrino experiment known as SAGE, a successful effort to measure the low-energy solar-neutrino flux. Today they are helping to build a next-generation solar-neutrino detector at the Sudbury Neutrino Observatory in Canada. The LSND collaboration is now planning an experiment at Fermilab, hoping to pin down the size of neutrino masses. And finally theorists at Los Alamos are continuing to study the role of neutrinos as drivers of supernovae and the possible role of neutrinos as components of the dark matter that binds together the large-scale structures in the universe.

Thus, the type of research that Reines and Cowan took on in the 1950s is very much a part of the Los Alamos tradition, held to this day, of doing fundamental science side by side with mission-oriented work. Their project was also the first in a series of extraordinary efforts at Los Alamos to chase down and uncover the properties of the elusive neutrinos. To honor Reines, Cowan, and their colleagues and to celebrate their discovery and the work it inspired, this issue is devoted to the neutrino and its remaining mysteries.

In closing, let us note that the Nobel Prize for the detection of the neutrino is not only a great honor for Reines personally, but also for Los Alamos National Laboratory—this is the first time Laboratory-sponsored work received such recognition. The Laboratory is extremely proud of Reines for bringing that honor to it. Very sadly, Cowan was not alive to share the award with Reines, but his equal contribution is recognized by all, and in that knowledge, portraits of both men are now hanging beside some of the other great scientists who have graced this institution. ∎

Neutrinos, they are very small.
> *They have no charge and have no mass*

And do not interact at all.
> *The earth is just a silly ball*

To them, through which they simply pass,
> *Like dustmaids down a drafty hall*

Or photons through a sheet of glass.
> *They snub the most exquisite gas,*

Ignore the most substantial wall,
> *Cold-shoulder steel and sounding brass,*

Insult the stallion in his stall,
> *And, scorning barriers of class,*

Infiltrate you and me! Like tall
> *And painless guillotines, they fall*

Down through our heads into the grass.
> *At night, they enter at Nepal*

And pierce the lover and his lass
> *From underneath the bed—you call*

*It wonderful; I call it crass**

1953-1956

The Reines-Cowan Experiments

Detecting the Poltergeist

A compilation of papers and notes by Fred Reines and Clyde Cowan, Jr.

Hanford Team 1953

Savannah Team 1955

Within the photograph, a sign reads:

City Limit-Santa Fe, N.M.

Access to this area is restricted to those who have either checked themselves with a hand-and-foot counter since entering the building or who are wearing white gloves and shoe covers.

The Hanford Team: (on facing page, left to right, back row) F. Newton Hayes, Captain W. A. Walker, T. J. White, Fred Reines, E. C. Anderson, Clyde Cowan, Jr., and Robert Schuch (inset); not all team members are pictured.
The Savannah River Team: (clockwise, from lower left foreground) Clyde Cowan, Jr., F. B. Harrison, Austin McGuire, Fred Reines, and Martin Warren; (left to right, front row) Richard Jones, Forrest Rice, and Herald Kruse.

In 1951, when Fred Reines first contemplated an experiment to detect the neutrino, this particle was still a poltergeist, a fleeting yet haunting ghost in the world of physical reality. All its properties had been deduced but only theoretically. Its role was to carry away the missing energy and angular momentum in nuclear beta decay, the most familiar and widespread manifestation of what is now called the weak force. The neutrino surely had to exist. But someone had to demonstrate its reality. The relentless quest that led to the detection of the neutrino started with an energy crisis in the very young field of nuclear physics.

The Missing Energy and the Neutrino Hypothesis

During the early decades of this century, when radioactivity was first being explored and the structure of the atomic nucleus unraveled, nuclear beta decay was observed to cause the transmutation of one element into another. In that process, a radioactive nucleus emits an electron (or a beta ray) and increases its positive charge by one unit to become the nucleus of another element. A familiar example is the beta decay of tritium, the heaviest isotope of hydrogen. When it undergoes beta decay, tritium emits an electron and turns into helium-3.

The process of beta decay was studied intensely. In particular, scientists measured the energy of the emitted electron. They knew that a definite amount of nuclear energy was released in each decay reaction and that, by the law of energy conservation, the released energy had to be shared by the recoil nucleus and the electron.

The requirements of energy conservation, combined with those of momentum conservation, implied that the electron should always carry away the same amount of energy (see the box "Beta Decay and the Missing Energy" on the facing page). That expectation seemed to be borne out in some experiments, but in 1914, to the great consternation of many, James Chadwick showed definitively that the electrons emitted in beta decay did not have one energy or even a discrete set of energies. Instead, they had a continuous spectrum of energies. Whenever the electron energy was at the maximum observed, the total energy before and after the reaction was the same, that is, energy was conserved. But in all other cases, some of the energy released in the decay process appeared to be lost.

In late 1930, Wolfgang Pauli endeavored to save the time-honored law of energy conservation by proposing what he himself considered a "desperate remedy" (see the box "The Desperate Remedy" on this page)—

The Desperate Remedy

4 December 1930
Gloriastr.
Zürich

Physical Institute of the
Federal Institute of Technology (ETH)
Zürich
Dear radioactive ladies and gentlemen,
   As the bearer of these lines, to whom I ask you to listen graciously, will explain more exactly, considering the 'false' statistics of N-14 and Li-6 nuclei, as well as the continuous β-spectrum, I have hit upon a desperate remedy to save the "exchange theorem"[*] of statistics and the energy theorem. Namely [there is] the possibility that there could exist in the nuclei electrically neutral particles that I wish to call neutrons,[**] which have spin 1/2 and obey the exclusion principle, and additionally differ from light quanta in that they do not travel with the velocity of light: The mass of the neutron must be of the same order of magnitude as the electron mass and, in any case, not larger than 0.01 proton mass. The continuous β-spectrum would then become understandable by the assumption that in β decay a neutron is emitted together with the electron, in such a way that the sum of the energies of neutron and electron is constant.
   Now, the next question is what forces act upon the neutrons. The most likely model for the neutron seems to me to be, on wave mechanical grounds (more details are known by the bearer of these lines), that the neutron at rest is a magnetic dipole of a certain moment μ. Experiment probably required that the ionizing effect of such a neutron should not be larger than that of a γ ray, and thus μ should probably not be larger than e.10^{-13} cm.
   But I don't feel secure enough to publish anything about this idea, so I first turn confidently to you, dear radioactives, with a question as to the situation concerning experimental proof of such a neutron, if it has something like about 10 times the penetrating capacity of a γ ray.
   I admit that my remedy may appear to have a small *a priori* probability because neutrons, if they exist, would probably have long ago been seen. However, only those who wager can win, and the seriousness of the situation of the continuous β-spectrum can be made clear by the saying of my honored predecessor in office, Mr. Debye, who told me a short while ago in Brussels, "One does best not to think about that at all, like the new taxes." Thus one should earnestly discuss every way of salvation.—So, dear radioactives, put it to test and set it right.—Unfortunately, I cannot personally appear in Tübingen, since I am indispensable here on account of a ball taking place in Zürich in the night from 6 to 7 of December.—With many greetings to you, also to Mr. Back, your devoted servant,

   W. Pauli

[*]In the 1957 lecture, Pauli explains, "This reads: exclusion principle (Fermi statistics) and half-integer spin for an odd number of particles; Bose statistics and integer spin for an even number of particles."

This letter, with the footnote above, was printed in the September 1978 issue of *Physics Today*.

_____

[**]Pauli originally called the new particle the neutron (or the "neutral one"). Later, Fermi renamed it the neutrino (or the "little neutral one").

Beta Decay and the Missing Energy

In all types of radioactive decay, a radioactive nucleus does not only emit alpha, beta, or gamma radiation, but it also converts mass into energy as it goes from one state of definite energy (or equivalent rest mass M1) to a state of lower energy (or smaller rest mass M2). To satisfy the law of energy conservation, the total energy before and after the reaction must remain constant, so the mass difference must appear as its energy equivalent (kinetic energy plus rest mass energy) among the reaction products.

Early observations of beta decay suggested that a nucleus decays from one state to a state with one additional unit of positive charge by emitting a single electron (a beta ray). The amount of energy released is typically several million electron volts (MeV), much greater than the rest mass energy of the electron (0.51 MeV). Now, if a nucleus at rest decays into two bodies—the final nucleus and the electron—the law of momentum conservation implies that the two must separate with equal and opposite momentum (see top illustration). Thus, conservation of energy and momentum implied that the electron from a given beta-decay process would be emitted with a constant energy.

Moreover, since a nucleus is thousands of times heavier than an electron, its recoil velocity would be negligible compared with that of the electron, and the constant electron energy would carry off just about all the energy released by the decay.

The graph (center) shows the unexpected results obtained from experiment. The electrons from beta decay were not emitted with a constant energy. Instead, they were emitted with a continuous spectrum of energies up to the expected value. In most instances, some of the energy released in the decay appeared to be lost. Scientists of the time wondered whether to abandon the law of energy conservation when considering nuclear processes.

Three-Body Decay and the Neutrino Hypothesis.

Pauli's solution to the energy crisis was to propose that the nucleus underwent beta decay and was transformed into three bodies: the final nucleus, the electron, and a new type of particle that was electrically neutral, at least as light as the electron, and very difficult to detect (see bottom illustration). Thus, the constant energy expected for the electron alone was really being shared between these two light particles, and the electron was being emitted with the observed spectrum of energies without violating the energy conservation law.

Pauli made his hypothesis in 1930, two years before Chadwick discovered the neutron, and he originally called the new particle the neutral one (or neutron). Later, when Fermi proposed his famous theory of beta decay (see the box "Fermi's Theory of Beta Decay and Neutrino Processes" on the next page), he renamed it the neutrino, which in Italian means the "little neutral one."

Two-Body Final State*

Helium-3 (1, 2)

Tritium (2, 1)

Recoil nucleus and electron separate with equal and opposite momentum.

e^-

$(N, Z) \rightarrow (N-1, Z+1) + e^-$,

where N = number of neutrons, and Z = number of protons.

Observed spectrum of energies

Expected electron energy

Number of electrons

Energy

Endpoint of spectrum

Three-Body Final State

Helium-3 (1, 2)

Tritium (2, 1)

Electron and neutrino share the available energy.

Electron   Antineutrino

$(N, Z) \rightarrow (N-1, Z+1) + e^- + \bar{\nu}$.

Fermi's Theory of Beta Decay and Neutrino Processes

In 1934, long before the neutrino was detected in an experiment, Fermi gave the neutrino a reality by writing down his simple and brilliant model for the beta decay process. This model has inspired the modern description of all weak-interaction processes. Fermi based his model on Dirac's quantum field theory of electromagnetism in which two electron currents, or moving electrons, exert force on each other through the exchange of photons (particles of light). The upper diagram represents the interaction between two electrons. The initial state of the system is on the left, and the final state is on the right. The straight arrows represent currents, or moving electrons, and the wiggly line between the currents represents the emission of a photon by one current and its absorption by another. This exchange of a photon causes the electrons to repel each other. Note that the photon has no mass, a fact related to the unlimited range of the electromagnetic force.

Electron Scattering
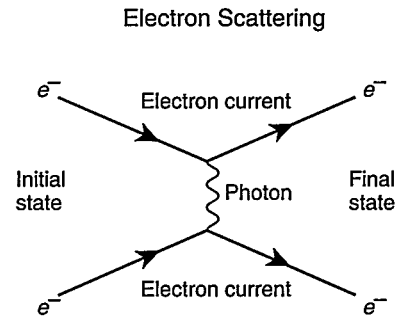
The fundamental process that takes place in beta decay (see lower diagram) is the change of a neutron into a proton, an electron, and an antineutrino. The neutron may be a free particle, or it may be bound inside the nucleus.

In analogy with quantum electrodynamics, Fermi represented beta decay as an interaction between two currents, each carrying the weak charge. The weak charge is related to the electric charge. Unlike the electromagnetic force, however, the weak force has a very short range. In Fermi's theory, the range of the force is zero, and the currents interact directly at a single point. The interaction causes a transfer of electric (weak) charge between the currents so that, for example, the neutron current gains one unit of charge and transforms into a proton current, while the electron current loses one unit of charge and transforms into a neutrino current.*

Neutron Beta Decay
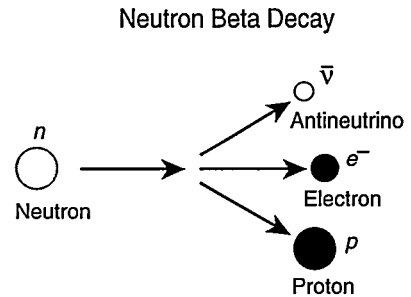
Because Fermi's theory is a relativistic quantum field theory, a single current-current interaction describes all weak-interaction processes involving the neutron, proton, electron, and neutrino or their antiparticles. As a result, we can represent all these weak-interaction processes with one basic diagram (on facing page, upper left corner).

---

*In the modern theory, the currents interact through the exchange of the W, a very heavy particle analogous to the photon. The W carries one unit of electric charge and one unit of weak isotopic charge between the weak currents.

a new subatomic particle that shares the available energy with the electron. To produce the observed energy spectrum, this new particle, later named the neutrino ("little neutral one"), could have a mass no larger than that of the electron. It had to have no electric charge. And like electrons and protons, the only subatomic particles known at that time, it had to be a fermion, a particle having half-integer spin (or intrinsic angular momentum). It would therefore obey

the Pauli exclusion principle according to which no two identical neutrinos can be in the same state at the same time. Once created, the neutrino would speed away from the site at, or close to, the speed of light. But Pauli was concerned that the neutrinos he had postulated should have been already detected.

Shortly thereafter, in a brilliant burst of insight, Enrico Fermi formulated a mathematical theory that involved the neutrino and that has endured with

little modification into the present. This theory postulates a force for beta decay and incorporates several brand-new concepts: Pauli's neutrino hypothesis, Dirac's ideas about the creation of particles, and Heisenberg's idea that the neutron and the proton were related to each other. In Fermi's theory of beta decay, this weak force, so called because it was manifestly much weaker than the electromagnetic force, turns a neutron into a proton and simultane-

Basic Current-Current Interaction
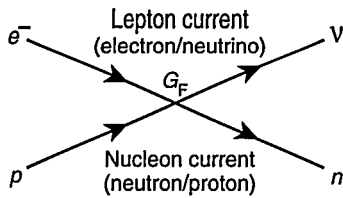
In analogy with the electric current, each weak current is depicted as a moving particle (straight arrow) carrying the weak charge. At the point where they interact, the two currents exchange one unit of electric (weak) charge.

One can adapt the basic diagram to each reaction by deciding which particles (or antiparticles) are to be viewed as the initial state and which as the final state. (Particles are represented by arrows pointing to the final state, whereas antiparticles point backward, to the initial state.) Since all the reactions described by the diagram stem from the same interaction, they have the same overall strength given by G_F, Fermi's constant. However, kinematic factors involving the amount and distribution of available energy and momentum in the initial and final states affect the overall reaction rate. Three reactions are illustrated in the lower diagrams.

In the first reaction, neutron beta decay (lower left), the neutron starts out alone, but the interaction of two currents is responsible for the decay. The neutron (current) turns into a proton, and the charge is picked up by the electron/neutrino (current) that creates a particle (electron) and an antiparticle (antineutrino). Note that the direction of the arrow for the neutrino points backwards, to the initial state, to indicate that an antineutrino has appeared in the final state.

In the second reaction, electron capture (lower center), the initial state is a proton (current) and an electron (current). The weak interaction between the two currents triggers the exchange of one unit of charge so that the proton turns into a neutron while the electron turns into a neutrino. The reverse process is also possible.

In the third case, inverse beta decay (lower right), the initial state is an antineutrino (current) and a proton (current). The weak interaction between the two currents triggers the exchange of one unit of charge so that the antineutrino turns into an antielectron (positron) while the proton turns into a neutron. Again, the arrows pointing backward indicate that an antineutrino in the initial state has transformed into an antielectron in the final state. The reverse process is also possible.

Neutron Beta Decay

$$n \rightarrow p + e^- + \bar{\nu}$$

Electron Capture

$$e^- + p \rightarrow n + \nu$$

Inverse Beta Decay

$$\bar{\nu} + p \rightarrow n + e^+$$
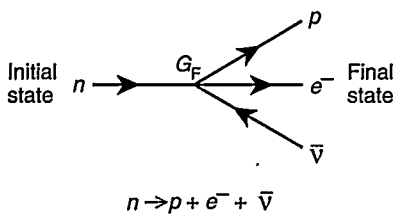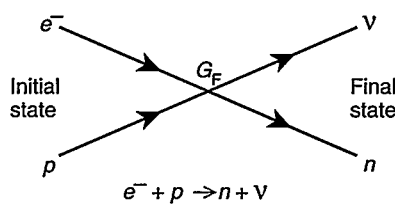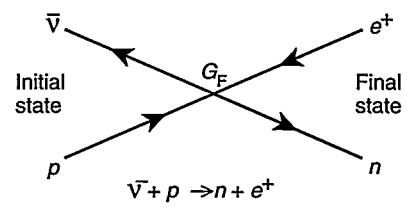
ously creates an electron and an anti-neutrino (see the box on this page). The force can act on a free neutron or on a neutron bound inside a nucleus.

Fermi's theory is remarkable in that it accounts for all the observed properties of beta decay. It correctly predicts the dependence of the radioactive-nucleus lifetime on the energy released in the decay. It also predicts the correct shape of the energy spectrum of the emitted electrons. Its success was taken

as convincing evidence that a neutrino is indeed created simultaneously with an electron every time a nucleus disintegrates through beta decay.

Almost as soon as the theory was formulated, Hans Bethe and Rudolf Peierls understood that Fermi's theory of the weak force suggested a reaction by which a free neutrino would interact with matter and be stopped. As Bethe and Bacher noted (1936),

"[I]t seems practically impossible to

detect neutrinos in the free state, i.e., after they have been emitted by the radioactive atom. There is only one process which neutrinos can certainly cause. That is the inverse beta process, consisting of the capture of a neutrino by a nucleus together with the emission of an electron (or positron)."

Unfortunately, the weak force is so weak that the probability of inverse beta decay was calculated to be close to zero. A target would have to be light-years

thick before it would have a good chance of stopping a neutrino. The possibility of detecting the neutrino seemed nil. But two things changed that prospect: first, the advent of very intense sources of neutrinos—fission bombs and fission reactors—and, second, the intense drive of a young man from New Jersey to make his mark in the world of fundamental physics.

Fred Reines and Los Alamos

Fred Reines had become interested in mathematics and physics while studying at the Stevens Institute of Technology, and during graduate studies at New York University, he wrote a Ph.D. thesis elaborating on Bohr's liquid-drop model of nuclear fission. In 1944, he joined the Manhattan Project at Los Alamos and became a member of the Theoretical Division.

During the late forties and early fifties, after the first atomic bomb had been built at Los Alamos, the Laboratory's mission was intensely focused on building a reliable stockpile of fission weapons and developing the thermonuclear bomb. Reines was in charge of several projects related to testing nuclear weapons in the Pacific. In retrospect, Reines explains (unpublished notes for a talk given at Los Alamos):

"Bomb testing was an exercise in thinking big, in the 'can do' spirit. In the George Shot, for example, the signal cables running from the shot tower to the instrumentation bunker had to be shielded from the enormous gamma-ray flux from the explosion; otherwise, that flux would generate a huge current surge in those cables that would destroy all our electronics. The only thing available for shielding on the scale we needed was the island itself. So we dug up one side of the island and put it on top of the other.

"That can do spirit permeated our thinking. Whenever we thought about new projects, the idea was to set the most interesting (and fundamental) goal without initial concern as to feasibility

or practical uses. We could count on the latest technology being available to us at Los Alamos as a result of the instrumentation needs of the weapons program, and that fact fed our confidence. To his credit, Norris Bradbury, the Director who took over after Oppenheimer, lent enormous support to surrounding the nuclear weapons effort at Los Alamos with a broad scientific and technological base."

The bomb-test steering and liaison group, in which Fred Reines participated, was interested in fundamental questions. New physics experiments that could be mounted as part of nuclear weapons tests were the topic of numerous free-ranging discussions in the group. It seemed appropriate that the unusually intense flux of thermal radiation, neutrons, and gamma rays produced by the bomb be used to study new phenomena.

The scientists in this group were even aware of the incredibly intense flux of antineutrinos produced when the fissioning, or splitting, of atomic nuclei during the neutron chain reaction gives rise to a host of unstable nuclei. The weak interactions then become important in changing the identity of those nuclei as they follow their decay paths to lower and lower energy states. Each fission event gives rise to an average of six beta-decay processes, each of which produces an antineutrino. Thus, those beta decays result in a short but intense burst of antineutrinos.

In 1951, Reines thought about using that intense burst in an experiment designed to detect the neutrino. He had returned from the very successful Greenhouse tests in Eniwetok Atoll, in the Pacific, and became captivated by the "impossible challenge" to detect the elusive free neutrino using neutrinos from the bomb. After having been involved for seven years in the weapons program, Reines asked J. Carson Mark, leader of the Theoretical Division, for some time to think about more fundamental questions.

The bomb was not only an intense neutrino source but also so short-lived

that the number of background events mimicking neutrino-induced events would be minimized. That summer, Reines mentioned his plan to Enrico Fermi and even described the need for what was then considered to be a very large scale detector. Reines estimated that a sensitive mass of about one ton would be needed to stop a few neutrinos. At the time, Reines did not know how to build such a large detector, and evidently, neither did Fermi. However, both Fermi and Hans Bethe thought that the bomb was the most promising neutrino source.

A few months later, Reines was able to interest one of his Los Alamos colleagues to participate in his quest. As Reines observed (unpublished notes), "It was my singular good fortune to be joined by Clyde L. Cowan, Jr., whom I had met in connection with Operation Greenhouse and who became my very stimulating and capable collaborator."

Cowan had studied chemical engineering as an undergraduate and, during World War II, was awarded the Bronze Star for his work on radar at the British Branch of the Radiation Laboratory of the Massachusetts Institute of Technology. His Ph.D. thesis at George Washington University was on the absorption of gamma radiation. In 1949, he joined Los Alamos Scientific Laboratory. Like Reines, he became heavily involved in the weapons testing program in the Pacific. In late 1951, Reines and Cowan began "Project Poltergeist," the first experiment in neutrino physics.

The Signal of the Poltergeist

What happens when neutrinos enter matter? Most of the time, they pass straight through without scattering, but Fermi's theory of the weak force predicts that the neutrino can induce an inversion of beta decay (see the box "Fermi's Theory of Beta Decay and Neutrino Processes" on page 8). In particular, the antineutrino (the antiparticle of the neutrino) will occasionally

interact with a nucleus through the weak force and will induce the transformation of a proton into a neutron. This inverse of the usual beta-decay process results in a nucleus with one less unit of positive charge. That charge is picked up by the antineutrino, which transforms into a positron:

$$\bar{\nu} + N\,(n, p) \rightarrow e^+ + N\,(n{+}1, p{-}1) \ ,$$
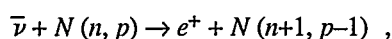
where n equals the number of neutrons and p equals the number of protons. If the nucleus happens to be that of hydrogen (a single proton), then the interaction produces a neutron and a positron:

$$\bar{\nu} + p \rightarrow n + e^+ \ .$$
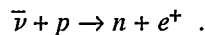
Reines and Cowan chose this latter reaction, the inverse beta decay on protons, to detect the free neutrino. The nuclear fission bomb would be their source of an intense flux of neutrinos (Figure 1). But they also needed to design a very large detector containing a sufficient number of target protons that would stop a few neutrinos. As Reines observed (unpublished notes),

"Our crude knowledge of the expected energy spectrum of neutrinos from a fission bomb suggested that the inverse beta decay reaction would occur several times in a several-ton detector located about 50 meters from the tower-based explosion of a 20-kiloton bomb. (Anyone untutored in the effects of nuclear explosions would be deterred by the challenge of conducting an experiment so close to the bomb, but we knew otherwise from experience and pressed on). The detector we dreamed up was a giant liquid scintillation device, which we dubbed 'El Monstro.' This was a daring extrapolation of experience with the newly born scintillation technique. The biggest detector until Cowan and I came along was only a liter or so in volume."

Their initial scheme was to use the newly discovered, liquid, organic scintillators as both the target for the neutrinos (these liquids had a high proportion
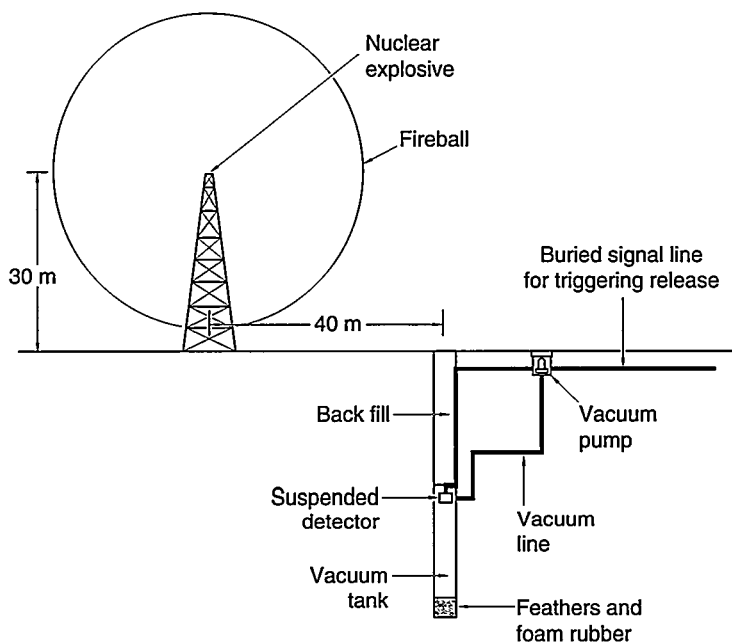
Figure 1. Detecting Neutrinos from a Nuclear Explosion
Antineutrinos from the fireball of a nuclear device would impinge on a liquid scintillation detector suspended in the hole dug below ground at a distance of about 40 meters from the 30-meter-high tower. In the original scheme of Reines and Cowan, the antineutrinos would induce inverse beta decay, and the detector would record the positrons produced in that process. This figure was redrawn courtesy of Smithsonian Institution.

of hydrogen) and the medium to detect the positron from inverse beta decay.

In 1950, several groups discovered that transparent organic liquids emit flashes of visible light when a charged particle or a gamma ray passes through them. These liquids had first been purified and then added to certain compounds. The light flashes are very weak but useful because their intensity is proportional to the energy of the charged particles or gammas. In a liquid scintillation counter, the light is collected by highly sensitive photomultiplier tubes located on the boundary of the detector. These phototubes convert light into electrical signals in proportion to the light intensity.

Figure 2 outlines the processes that would convert the energy of a positron from inverse beta decay into a measurable signal. The first small liquid-scintillation counters had already been developed, and one of those initial developers, F. B. (Kiko) Harrison, was at Los Alamos.

Wright Langham, leader of the Health Division's research group, had recruited Harrison to help design such counters for measuring radiation in biological samples. Harrison was one of the designers of the prompt-coincidence technique (see the section "The First Large Detector" on page 14) to distinguish spurious noise in the photomultiplier tubes from the signals generated by light flashes.

Once the idea for a new detector had been shaped, Reines and Cowan developed an audacious design for their experiments (shown in Figure 1). As Cowan (1964) vividly described it,

"We would dig a shaft near 'ground zero' about 10 feet in diameter and about 150 feet deep. We would put a tank, 10 feet in diameter and 75 feet long on end at the bottom of the shaft. We would then suspend our detector from the top of the tank, along with its recording apparatus, and back-fill the shaft above the tank.

"As the time for the explosion
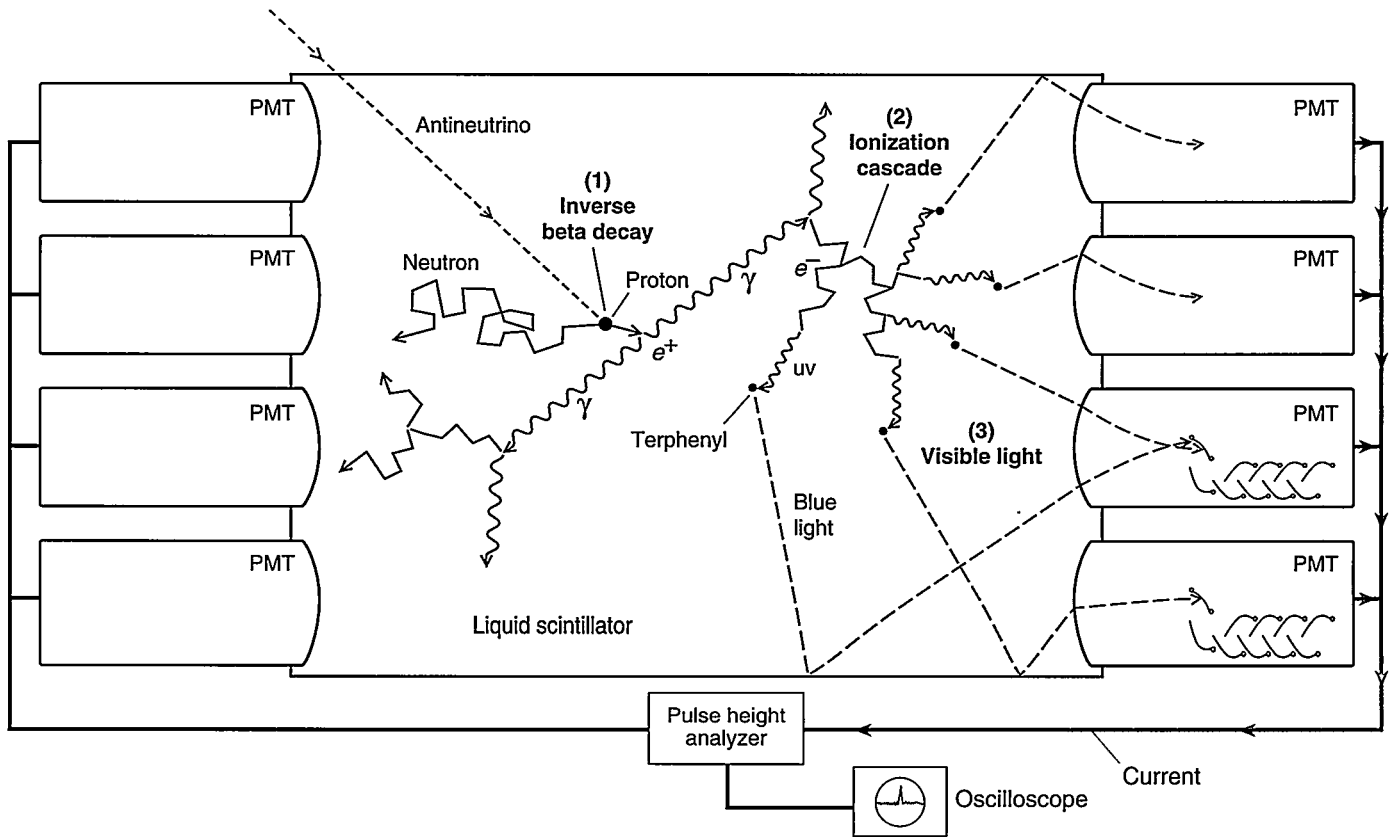
Figure 2. Liquid Scintillation Counter for Detecting the Positron from Inverse Beta Decay

Reines and Cowan planned to build a counter filled with liquid scintillator and lined with photomultiplier tubes (PMTs), the "eyes" that would detect the positron from inverse beta decay, which is the signal of a neutrino-induced event. The figure illustrates how the liquid scintillator converts a fraction of the energy of the positron into a tiny flash of light. The light is shown traveling through the highly transparent liquid scintillator to the PMTs, where the photons are converted into an electronic pulse that signals the presence of the positron. Inverse beta decay (1) begins when an antineutrino (red dashed line) interacts with one of the billions and billions of protons (hydrogen nuclei) in the molecules of the liquid. The weak charge-changing interaction between the

antineutrino and the proton causes the proton to turn into a neutron and the antineutrino to turn into a positron (e^+). The neutron wanders about undetected. The positron, however, soon collides with an electron (e^-), and the particle-antiparticle pair annihilates into two gamma rays (γ) that travel in opposite directions. Each gamma ray loses about half its energy each time it scatters from an electron (Compton scattering). The resulting energetic electrons scatter from other electrons and radiate photons to create an ionization cascade (2) that quickly produces large numbers of ultraviolet (uv) photons.
The scintillator is a highly transparent liquid (toluene) purposely doped with terphenyl. When it becomes excited by absorbing the uv photons, it scintillates

by emitting visible photons as it returns to the ground (lowest-energy) state (3). Because the liquid scintillator is transparent to visible light, about 20 percent of the visible photons are collected by the PMTs lining the walls of the scintillation counter. The rest are absorbed during the many reflections from the counter walls. A visible photon releases an electron from the cathode of a phototube. That electron then initiates the release of further electrons from each dynode of the PMT, a process resulting in a measurable electrical pulse. The pulses from all the tubes are combined, counted, processed, and displayed on an oscilloscope screen.

approached, we would start vacuum pumps and evacuate the tank as highly as possible. Then, when the countdown reached 'zero,' we would break the suspension with a small explosive, allowing the detector to fall freely in the

vacuum. For about 2 seconds, the falling detector would be seeing the antineutrinos and recording the pulses from them while the earth shock [from the blast] passed harmlessly by, rattling the tank mightily but not disturbing our falling

detector. When all was relatively quiet, the detector would reach the bottom of the tank, landing on a thick pile of foam rubber and feathers.

"We would return to the site of the shaft in a few days (when the

surface radioactivity had died away sufficiently) and dig down to the tank, recover the detector, and learn the truth about neutrinos!"

This extraordinary plan was actually granted approval by Laboratory Director Norris Bradbury. Although the experiment would only be sensitive to neutrino cross sections of 10^{-40} square centimeters, 4 orders of magnitude larger than the theoretical value, Bradbury was impressed that the plan was sensitive to a cross section 3 orders of magnitude smaller than the existing upper limit.[1] As Reines explains in retrospect (unpublished notes for a talk given at Los Alamos),

"Life was much simpler in those days—no lengthy proposals or complex review committees. It may have been that the success of Operation Greenhouse, coupled with the blessing given our idea by Fermi and Bethe, eased the path somewhat!"

As soon as Bradbury approved the plan, work started on building and testing El Monstro. This giant liquid-scintillation device was a bipyramidal tank about one cubic meter in volume. Four phototubes were mounted on each of the opposing apexes, and the tank was filled with very pure toluene activated with terphenyl so that it would scintillate. Tests with radioactive sources of electrons and gamma rays proved that it was possible to "see" into a detector of almost any size.

Reines and Cowan also began to consider problems associated with scaling up the detector. At the same time, work was proceeding on drilling the hole that would house the experiment at the Nevada Test Site and on designing the great vacuum tank

---

[1]H. R. Crane (1948) deduced the upper limit of 10^{-37} square centimeters on the cross sections for neutrino-induced ionization and inverse beta decay. This upper limit was based on null results from various small-scale experiments attempting to measure the results of neutrino absorption and from a theoretical limit deduced from the maximum amount of solar neutrino heating that could take place in the earth's interior and still agree with geophysical observations of the energy flowing out of the earth.
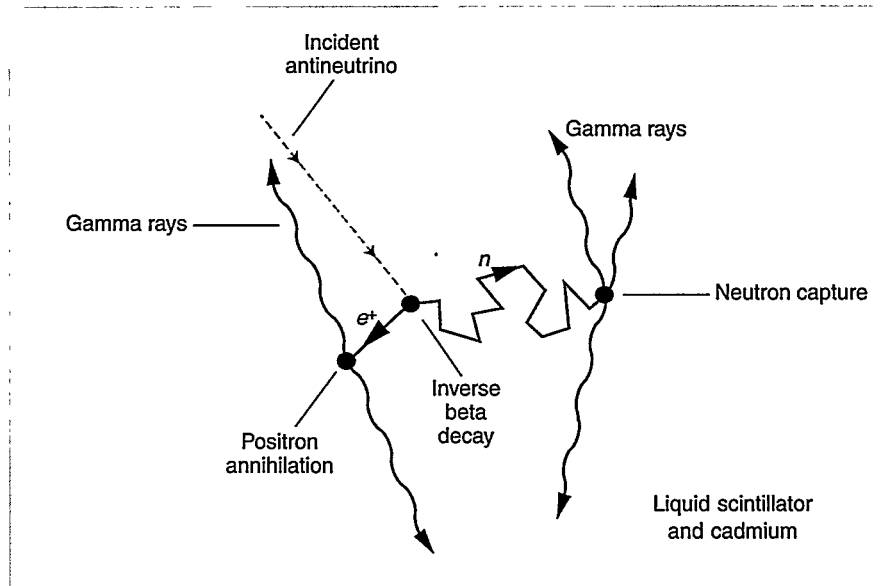
Figure 3. The Double Signature of Inverse Beta Decay
The new idea for detecting the neutrino was to detect both products of inverse beta decay, a reaction in which an incident antineutrino (red dashed line) interacts with a proton through the weak force. The antineutrino turns into a positron (e^+), and the proton turns into a neutron (n). In the figure above, this reaction is shown to take place in a liquid scintillator. The short, solid red arrow indicates that, shortly after it has been created, the positron encounters an electron, and the particle and antiparticle annihilate each other. Because energy has to be conserved, two gamma rays are emitted that travel in opposite directions and will cause the liquid scintillator to produce a flash of visible light. In the meantime, the neutron wanders about following a random path (longer, solid red arrow) until it is captured by a cadmium nucleus. The resulting nucleus releases about 9 MeV of energy in gamma rays that will again cause the liquid to produce a tiny flash of visible light. This sequence of two flashes of light separated by a few microseconds is the double signature of inverse beta decay and confirms the presence of a neutrino.

and its release mechanism.

But one late evening in the fall of 1952, immediately after Reines and Cowan had presented their plans at a Physics Division seminar, a new idea was born that would dramatically change the course of the experiment. J. M. B. Kellogg, leader of the Physics Division, had urged Reines and Cowan to review once more the possibility of using the neutrinos from a fission reactor rather than those from a nuclear explosion.

The neutrino flux from an explosion would be thousands of times larger than that from the most powerful reactor. The available shielding, however, would make the background noise from neutrons and gamma rays about the

same in both cases. Clearly, the nuclear explosion was the best available approach—unless the background could somehow be further reduced.

Suddenly, Reines and Cowan realized how to do it. The original plan had been to detect the positron emitted in inverse beta decay (see Figure 2), a process in which the weak interaction causes the antineutrino to turn into a positron and the proton to turn into a neutron. Being an antielectron, the positron would quickly collide with an electron, and the two would annihilate each other as they turned into pure energy in the form of two gamma rays traveling in opposite directions. Each gamma ray would have an energy equivalent to the rest mass of the

electron, namely, 0.51 million electron volt (MeV). The two gamma rays would accelerate electrons through Compton scattering and initiate a cascade of electrons that would eventually cause the liquid to scintillate. The tiny flash of visible light, efficiently converted into an electronic pulse, would be the signal of the positron.

The new idea was to detect not only the positron but also the neutron (see Figure 3). Once produced, the neutron bounces around and slows down as it collides with protons. It can be captured by a proton to produce deuterium, or heavy hydrogen. But if a nucleus such as cadmium is present, the neutron has a much greater chance of being captured. Adding a cadmium salt to the organic scintillator dramatically increases the cross section for absorbing (low-energy) neutrons. The capture process releases about 9 MeV of energy in gamma rays.

The average time between the flash of light from the positron-electron annihilation and that from the neutron capture is a few microseconds. Electronic circuits could be designed to detect this "delayed-coincidence" signature, two flashes of light (each within a well-defined energy range) separated by microseconds, and provide a powerful means to discriminate the signature of inverse beta decay from background noise. Thus, using the much smaller flux of reactor neutrinos became feasible. As Cowan (1964) remembers,

"Instead of detecting a burst of neutrinos in a second or two coming from the fury of a nuclear explosion, we would now be able to watch patiently near a reactor and catch one every few hours or so. And there are many hours available for watching in a month—or a year."

The First Large Detector

The group spent that winter building the detectors, developing various liquid-scintillator compositions, and testing the response of the detectors to gamma rays. Each detector

was about 28 inches in diameter and 30 inches high (see photo on this page), and 90 photomultiplier tubes penetrated its curved walls.

The phototubes were connected in two interleaved arrays, each of which would produce an electrical pulse in response to a light signal in the detector. The two pulses would then be sent to a prompt-coincidence circuit, which would accept them as a bona fide signal
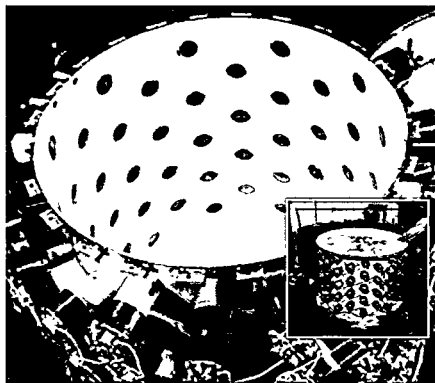
The Hanford Neutrino Detector
The background photo is a top view of the neutrino detector used in the Hanford experiments. It shows the interior of the 10-cubic-foot vat for the liquid scintillator and the 90 photomultiplier tubes, each with a 2-inch-diameter face that had a thin, photosensitive surface. The inset is a side view of the detector. Having a 300-liter capacity, "Herr Auge" (German for Mr. Eye, as this detector was named) · was the largest detector at the time.

only if they arrived simultaneously. That prompt-coincidence requirement helped eliminate counting the spurious dark current that arose spontaneously and at random in the phototubes themselves.

The team worked in an isolated, unheated building. Cowan (1964) reports how "some of our group swept the snow away from outside the building and set about casting many large blocks of paraffin wax and borax for use as neutron shielding when we would go to a reactor. Others began mixing gallons of liquid scintillator in batches with varying composition."

They had to use electrical heaters to

keep the toluene scintillator warm; otherwise, it would turn from transparent to cloudy. Soon, they discovered that one of the brands of mineral oil carried by a local druggist, when mixed with suitable chemicals, could serve as another liquid scintillator. Having a hydrogen density different from that of toluene, the mineral oil would yield a different measured rate for inverse beta decay and thus provide a consistency check on the experiment—of course, if the experimental error could be made small enough to make the difference visible.

The threesome who carried the primary responsibility for developing and testing the detector were F. Newton Hayes, Robert Schuch, and Ernest C. Anderson from Wright Langham's biomedical/health physics research group. Using various radioactive gamma-ray sources, they discovered that their large-volume liquid scintillation detectors were extremely efficient at detecting gamma rays, enough to revolutionize the counting of small amounts of radioactivity in bulk samples. The group realized they could test the radioactive content of the materials used to construct the detector and eliminate those that would add unduly to the background.

As Cowan (1964) reports, "We built a cylindrical well into one of the detectors and proceeded to put quantities of steel, liquids, wax, and other materials into it for testing. We found that brass and aluminum were quite radioactive compared to iron and steel, and that the potassium in the glass envelopes of our photomultiplier tubes would contribute to the detector backgrounds.

"During this time, one of our group, Robert Schuch, proposed making the well in the detector a bit larger so that we might be able to put a human being into the detector. This was done, and a number of people, including our secretary, were trussed up and lowered into the 18-inch hole. We found quite a detectable counting rate from everyone. It was due to the radioactive potassium-40 naturally present in the body."

The Whole-Body Counter
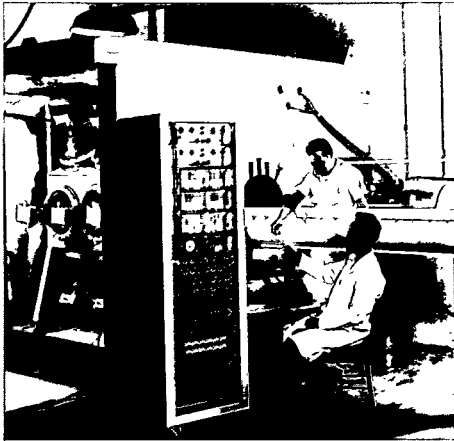
In 1956, Ernest C. Anderson, Robert Schuch, James Perrings, and Wright Langham developed the whole-body counter known as HUMCO I. Its design was a direct spinoff from the development of the first large liquid-scintillation detector used in Reines and Cowan's neutrino experiments at Hanford. HUMCO I measured low levels of naturally occurring radioactivity in humans. Later, it was used in a worldwide effort to determine the degree to which radioactive fallout from nuclear tests and other nuclear and natural sources was absorbed by the human body. The detector consisted of a cylindrical container filled with 140 gallons of liquid scintillator and surrounded by 108 photomultiplier tubes. The person being measured was placed in a slide and drawn into the detector. Gamma rays emitted by the naturally occurring radioisotope potassium-40 or the fallout isotope cesium-137, for example, would largely penetrate the detector's inner wall, excite the scintillator, and be detected. HUMCO II, which superseded HUMCO I in 1962, was nearly 10 times more sensitive, and its measurements were that much safer and quicker.

The top photo shows Anderson sitting at the controls of HUMCO II. To his right is the slide that would carry Schuch inside the detector for radioactive measurement.

In 1958, the human counter was demonstrated at the Atoms for Peace Conference held in Geneva. Built especially for this conference, the vertical counter was open on one side to allow a person to step in for measurement of internal radioactivity. The middle picture shows a conference participant getting ready to enter the detector under Newton Hayes' supervision.

The lower picture and diagram show the first human-radioactivity measurements carried out in the detector that served as the basis for HUMCO. The original purpose of that detector had been different: to determine the degree to which the natural gamma-ray activity of the materials used to shield the Hanford neutrino detector would add "noise" to the experiments. Schuch suggested that a larger insert into the detector would allow a small person to be placed inside and then be measured for gamma-ray activity. Langham, shown crouched inside the detector, was the only member of the team slim enough to fit in the narrow space.
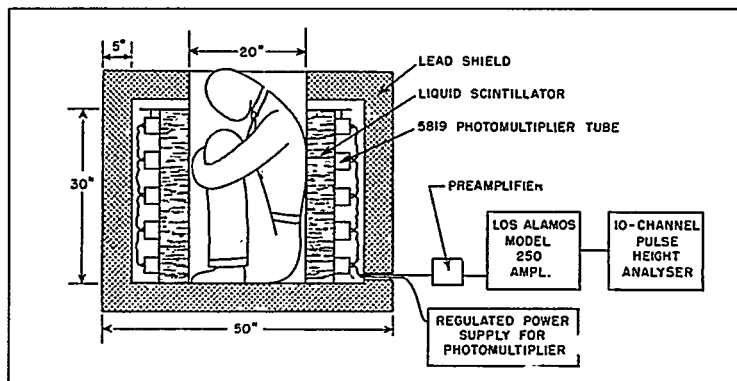
The Hanford Experiment
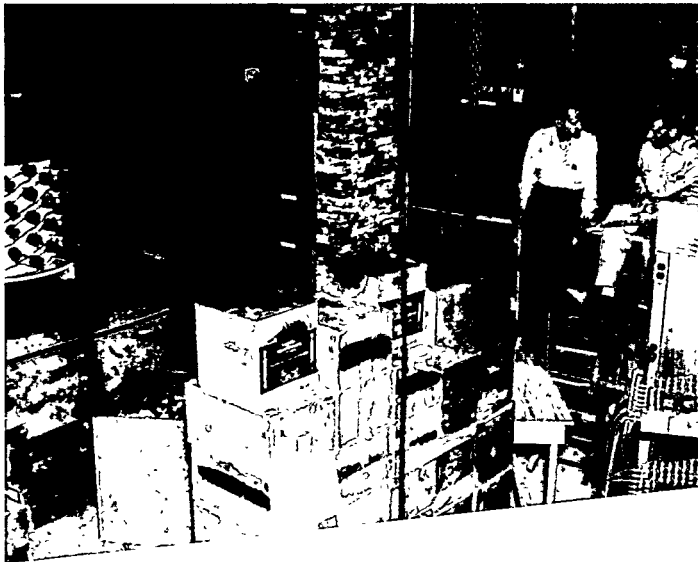
(a)

(b)

HANFORD CAFETERIA 1953

********** MENU **********

Todays Special: HOT TOLUENE FLUSH (Delicately Scented)

ENTREES

No. 1 TRIETHYL-BENZENE with PPO and ALPHA NPO.
       (Served with Methyl Borate by request only.)

No. 2 TERPHENYL. Served with hot toluene.

No. 3 MINERAL OIL RASH. Includes Toluene. PPO, Alpha NPO,
       Gasket Material and Pipe Dope.

No. 4 CADMIUM PROPIONATE COCKTAIL. Mixed with Methanol,
       "For that extra lift."

DESSERT

Green Men Cocktail.

Silica Gel.

Due to low humidity
requirements, all
orders will be served
without water.

You are cordially
invited to visit our
Armstrong sealed
kitchen.

F. N. Hayes and R.L. Schuch Prop.

(c)
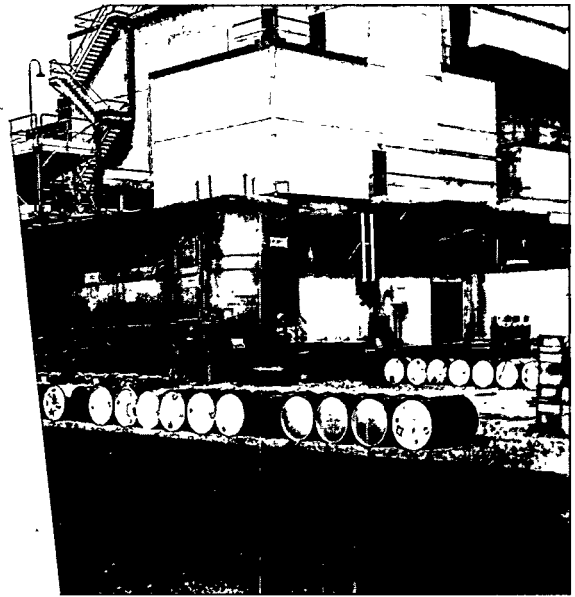
(d)

Amid the jumble of boxes and barrels, Los Alamos researchers were feverishly preparing for the Hanford experiment.

(a) F. Newton Hayes (left) and Clyde Cowan, Jr., discuss the search for the neutrino, while two workers (b) are shielding the face of the reactor to minimize the occurrence of background events. The top of Herr Auge, the neutrino detector, is shown surrounded by an incomplete shield made of boron-paraffin boxes and huge amounts of lead.

1953

(e)

(f)

(g)

Work was exciting, exhausting, all-consuming. But there was always time for fun. In the menu composed by Hayes and Robert Schuch (c), silica gel, the chemical "jello," is offered as a tongue-in-cheek dessert together with green men cocktail, a reminder of the green-colored solution left from rinsing the whole system before the experiment could start. The chemicals listed on the menu are some of the actual ingredients used in preparing the liquid scintillators that would fill the detector. The barrels (d) were filled with scintillator solution after the chemicals had carefully been weighed with the scale pictured in (e). Hayes is filling empty barrels (f) with that solution. The barrels would then be hauled onto the storage truck. Schuch is connecting pipes to the storage truck (g) in preparation for transferring the liquid scintillator into the mixing trailer. The two rows of valves and pipes were inside the mixing trailer (h). Through these pipes and the supply lines (i), the scintillator solution would flow into the detector.

These photos are from Robert Schuch's private collection.
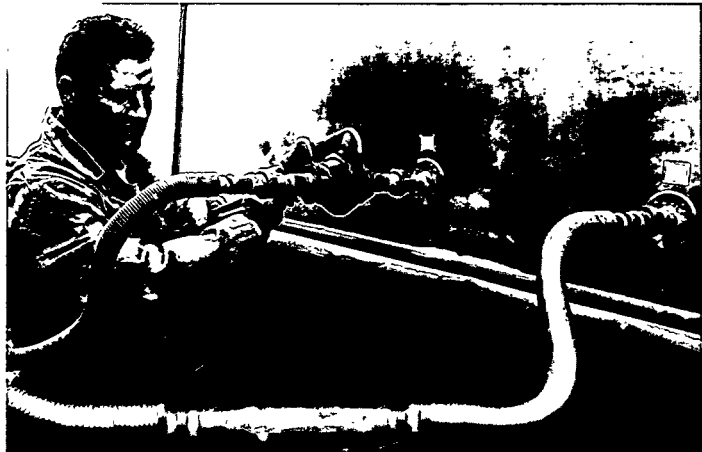
(h)

(i)

Schuch's idea gave birth to the Los Alamos total-immersion, or "whole-body," counter (see box "The Whole-Body Counter" on page 15), which was similar in design to the detector for Project Poltergeist but was built especially to count the radioactive contents of people. Since counting with this new device took only a few minutes, it was a great advance over the standard practice of using multiple Geiger counters or sodium iodide (NaI) crystal spectrometers in an underground laboratory. The Los Alamos whole-body counter was used during the 1950s to determine the degree to which radioactive fallout from nuclear tests and other nuclear and natural sources was taken up by the human body.

The Hanford Experiment

In the very early spring of 1953, the Project Poltergeist team packed up Herr Auge, the 300-liter neutrino detector, as well as numerous electronics and barrels of liquid scintillator, and set out for the new plutonium-producing reactor at the Hanford Engineering Works in Hanford, Washington. It was the country's latest and largest fission reactor and would therefore produce the largest flux of antineutrinos. Various aspects of the setup at Hanford are shown in the photo collage.

The equipment for the liquid scintillator occupied two trucks parked outside the reactor building. One was used to house barrels of liquid; in a second smaller truck, liquid scintillators were mixed according to various recipes before they would be pumped into the detector. Herr Auge was placed inside the reactor building, very near the face of the reactor wall, and was surrounded by the homemade boron-paraffin shielding intermixed with nearly all the lead shielding available at Hanford. This shield was to stop reactor neutrons and gamma rays from entering the detector and producing unwanted background. In all, 4 to 6 feet of paraffin alternated with 4 to 8 inches of lead.

The electronic gear for detecting the telltale delayed-coincidence signal from inverse beta decay was inside the reactor building. Its essential elements were two independent electronic gates: one to accept pulses characteristic of the positron signal and the other to accept pulses characteristic of the neutron-capture signal. The two circuits were connected by a time-delay analyzer.

If a pulse appeared in the output of the neutron circuit within 9 microseconds of a pulse in the output of the positron circuit, the count was registered in the channel that recorded delayed coincidences. Allowing for detector efficiencies and electronic gate settings and taking into account the neutrino flux from the reactor, the expected rate for delayed coincidences from neutrino-induced events was 0.1 to 0.3 count per minute.

For several months, the team stacked and restacked the shielding and used various recipes for the liquid scintillator (see Hanford Menu in "The Hanford Experiment" collage). Then they would set the electronics and listen for the characteristic double clicks that would accompany detection of the inverse beta decay. Despite the exhausting work, the results were not definitive. The delayed-coincidence background, present whether or not the reactor was on, was about 5 counts per minute, many times higher than the expected signal rate.

The scientists guessed that the background was due to cosmic rays entering the detector, but the addition of various types of shielding left the background rate unchanged. Subsequent work underground suggested that the Hanford background of delayed-coincidence pulses was indeed due to cosmic rays. Reines and Cowan (1953) reported a small increase in the number of delayed coincidences when the reactor was on versus when it was off. Furthermore, the increase was consistent with the number expected from the estimated flux of reactor neutrinos. This was tantalizing but insufficient evidence that neutrino

events were being detected. The Hanford experience was poignantly summarized by Cowan (1964).

"The lesson of the work was clear: It is easy to shield out the noise men make, but impossible to shut out the cosmos. Neutrons and gamma rays from the reactor, which we had feared most, were stopped in our thick walls of paraffin, borax and lead, but the cosmic ray mesons penetrated gleefully, generating backgrounds in our equipment as they passed or stopped in it. We did record neutrino-like signals but the cosmic rays with their neutron secondaries generated in our shields were 10 times more abundant than were the neutrino signals. We felt we had the neutrino by the coattails, but our evidence would not stand up in court."

The Savannah River Experiment

After the Hanford experience, the Laboratory encouraged Reines and Cowan to set up a formal group with the sole purpose of tracking neutrinos. Other than the scientists who had already been working on neutrinos, Kiko Harrison, Austin McGuire, and Herald Kruse (a graduate student at the time) were included in this group.

They spent the following year redesigning the experiment from top to bottom: detector, electronics, scintillator liquids, the whole works. The detector was entirely reconfigured to better differentiate between events induced by cosmic rays and those initiated in the detector by reactor neutrinos. Figure 4 shows the new design.

Two large, flat plastic tanks (called the "target tanks" and labeled A and B) were filled with water. The protons in the water provided the target for inverse beta decay; cadmium chloride dissolved in the water provided the cadmium nuclei that would capture the neutrons. The target tanks were sandwiched between three large scintillation detectors labeled I, II, and III (total capacity 4,200 liters), each

having 110 photomultiplier tubes to collect scintillation light and produce. electronic signals.

In this sandwich configuration, a neutrino-induced event in, say, tank A would create two pairs of proton prompt-coincidence pulses from detectors I and II flanking tank A. The first pair of pulses would be from positron annihilation and the second from neutron capture. The two pairs would be separated by about 3 to 10 microseconds. Finally, no signal would emanate from detector III because the gamma rays from positron annihilation and neutron capture in tank A are too low in energy to reach detector III.

Thus, the spatial origin of the event could be deduced with certainty, and the signals would be distinguished from false delayed-coincidence signals induced by stray neutrons, gamma rays, and other stray particles from cosmic-ray showers or from the reactor. These spurious signals would most likely trigger detectors I, II, and III in a random combination. The all-important electronics were designed primarily by Kiko Harrison and Austin McGuire.

The box entitled "Delayed-Coincidence Signals from Inverse Beta Decay" (page 22) illustrates delayed-coincidence signals from the detector's top triad (composed of target tank A and scintillation detectors I and II). Once the delayed-coincidence signals have been recorded, the neutrino-induced event is complete. The signals from the positron and neutron circuits, which have been stored on delay lines, are presented to the oscilloscopes.

Figure 5 shows a few samples of oscilloscope pictures—some are acceptable signals of inverse beta decay while others are not.

Austin McGuire was in charge of the design and construction of the "tank farm" that would house and transport the thousands of gallons of liquid scintillator needed for the experiment. Three steel tanks were placed on a flat trailer bed. The interior surfaces of the tanks were coated with epoxy to preserve the purity of the liquids.
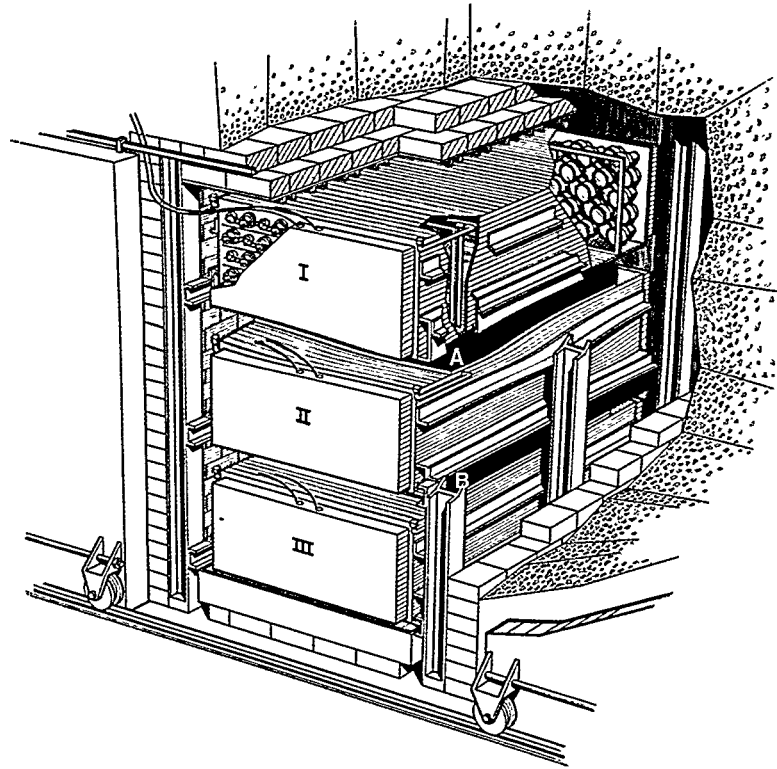
Figure 4. The Savannah River Neutrino Detector—A New Design
The neutrino detector is illustrated here inside its lead shield. Each of two large, flat plastic tanks (pictured in light blue and labeled A and B) was filled with 200 liters of water. The protons in the water provided the target for inverse beta decay; cadmium chloride dissolved in the water provided the cadmium nuclei that would capture the neutrons. The target tanks were sandwiched between three scintillation detectors (I, II, and III). Each detector contained 1,400 liters of liquid scintillator that was viewed by 110 photomultiplier tubes. Without its shield, the assembled detector weighed about 10 tons.

Today, the need for purity and cleanliness is becoming legendary as researchers build an enormous tank for the next generation of solar-neutrino experiments (see the article "Exorcising Ghosts" on page 136), but even in the 1950s, possible background contamination was an overriding concern.

Since the scintillator had to be kept at a temperature not lower than 60 degrees Fahrenheit, the outside walls of the tanks were wrapped with several layers of fiberglass insulating material, and long strips of electrical heating elements were embedded in the exterior insulation.

During the previous winter, while the equipment was being designed and built, John Wheeler encouraged and supported the team, and he helped

pave the way for the next neutrino measurement to be done at the new, very powerful fission reactor at the Savannah River Plant in South Carolina. By November 1955, the Los Alamos group was ready and once again packed up for the long trip to the Savannah River Plant.

The only suitable place for the experiments was a small, open area in the basement of the reactor building, barely large enough to house the detector. There, 11 meters of concrete would separate the detector from the reactor core and serve as a shield from reactor-produced neutrons, and 12 meters of overburden would help eliminate the troublesome background neutrons, charged particles, and gamma rays produced by cosmic rays.

The Savannah River Experiment
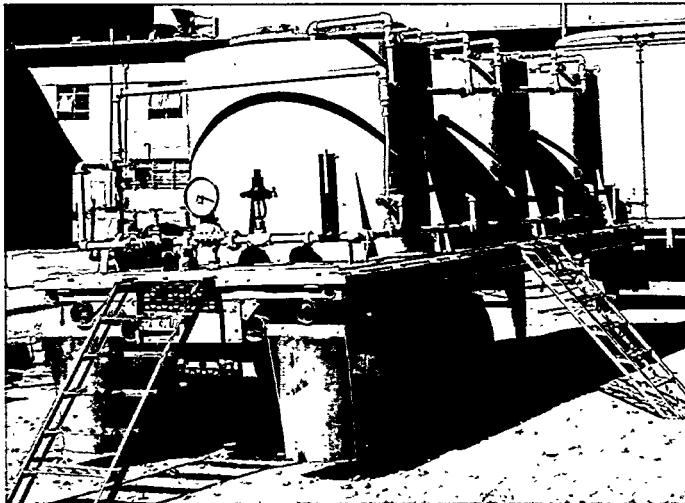
(a)
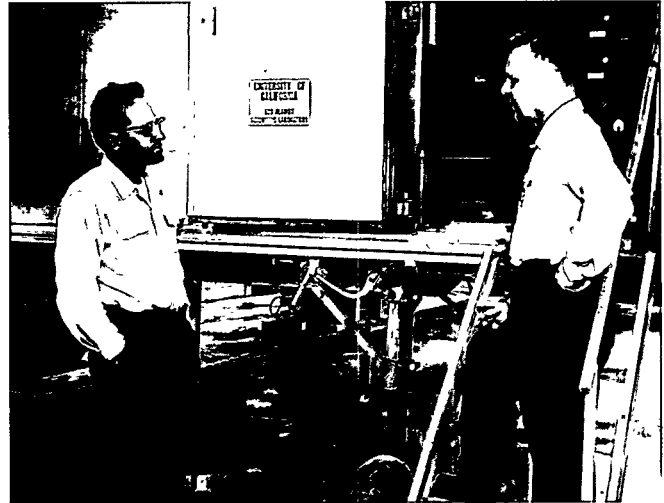
(b)
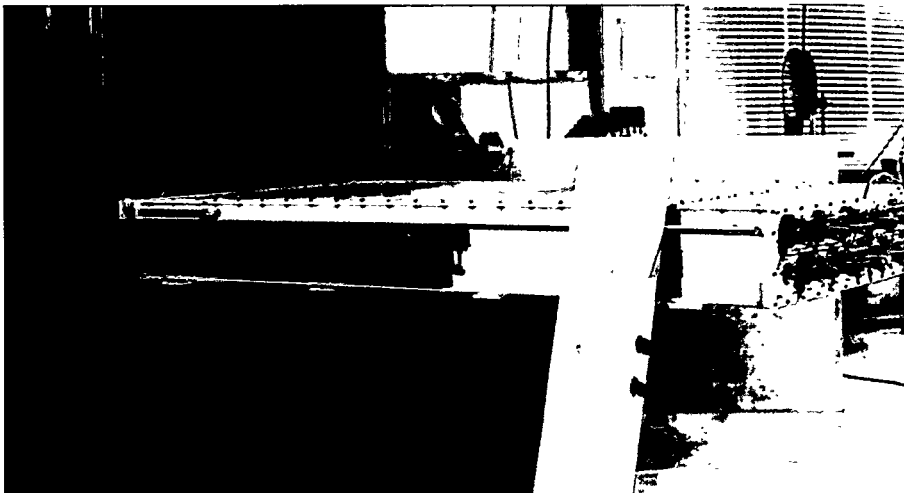
(c)
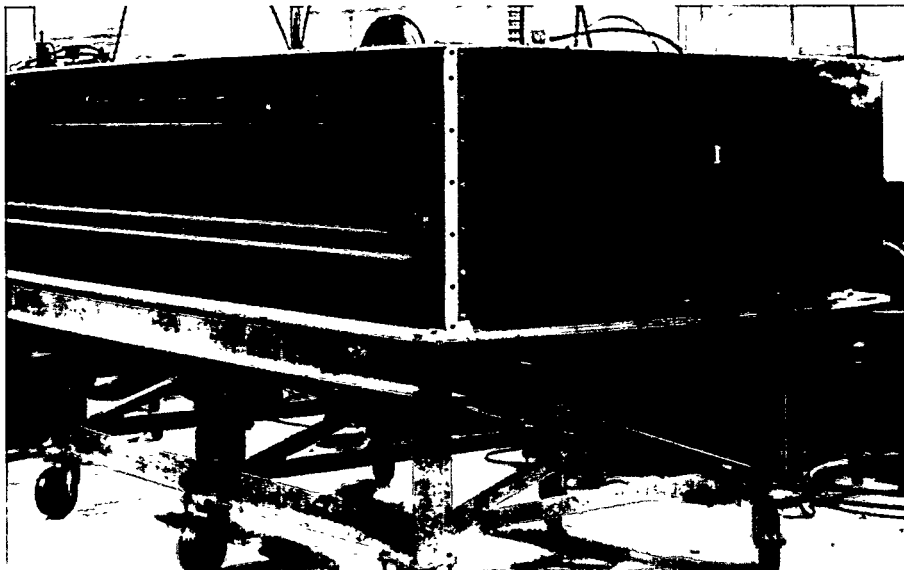
(d)

After years of intense work, the members of the Los Alamos team were ready for the Savannah River experiment that would fulfill their much expected goal—the definitive detection of the neutrino.

Pictured in (a) is the tank farm, which was composed of three 4,500-liter steel tanks placed on a flat-bed trailer. The liquid scintillator was stored and shipped in those tanks. The outside walls of the tanks were wrapped with fiberglass insulation, and long electrical heating strips were embedded in the insulation to prevent the temperature inside the tanks from falling below 60 degrees Fahrenheit. Had the temperature fallen below this limit, the liquid scintillator would have turned from transparent to cloudy and would have become unusable in the experiment. (b) Fred Reines (left) and Clyde Cowan, Jr., discuss their last-minute plans for the Savannah River experiment. No detail is left uncovered. Resting in a special forklift built to handle the detector sections, one of the two target tanks filled with water and cadmium chloride is shown (c) awaiting its assembly in the detector shield. A completed detector tank (d) is ready to be inserted into the shield. This tank was made of steel plate, but its bottom was a

(e)

cellular aluminum structure that would provide not only strength against bending but also little obstruction to the entry of gamma rays from below. (e) Pictured here is the additional shielding that surrounded the detector and allowed the team to test whether the signal was coming from background neutrons and gamma rays from the reactor. This makeshift shielding, which was 4 feet thick all around the detector, consisted of bags of sawdust soaked in water for increased density (the mean density was 0.5). Its effect was to decrease the reactor-associated accidental events, whereas the signal remained constant. (f) Los Alamos team members Richard Jones (left) and Martin Warren use a forklift to insert the top target tank into the detector shield. Moving by hydraulic control, heavy lead doors (pictured behind Warren) would enclose the detector when it was on. Preamplifiers placed on a rack (pictured behind Jones) boosted the small-voltage pulses from the photomultiplier tubes and sent them through coaxial cable to the electronics housed in a truck (g) that was parked outside the reactor building.

Photos (c), (d), (e), and (f) were reprinted courtesy of Smithsonian Institution.

(f)
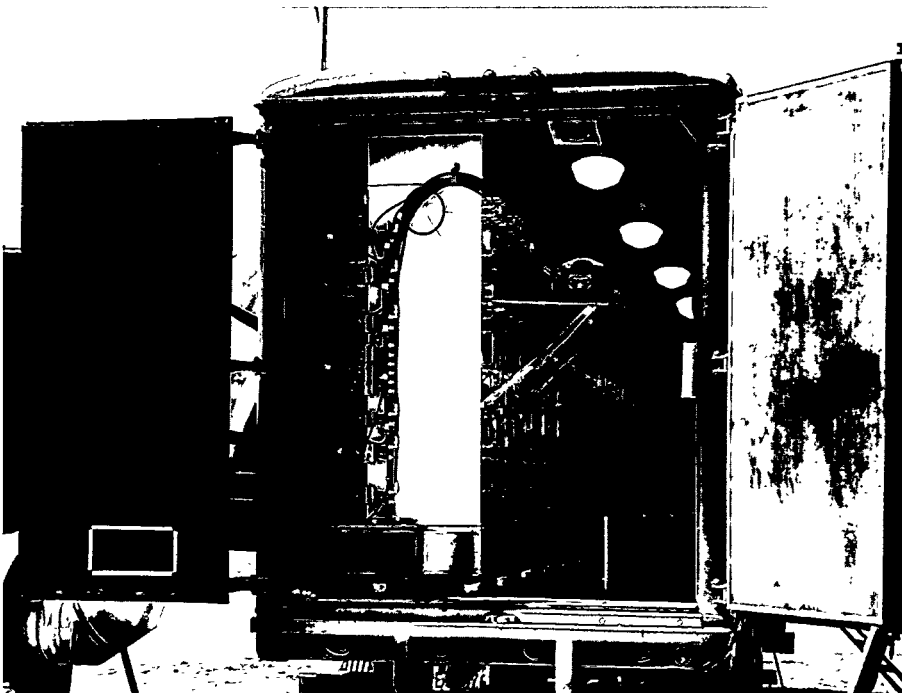
(g)

Delayed-Coincidence Signals from Inverse Beta Decay

(a) T = 0   Positron annihilation produces electron signal.
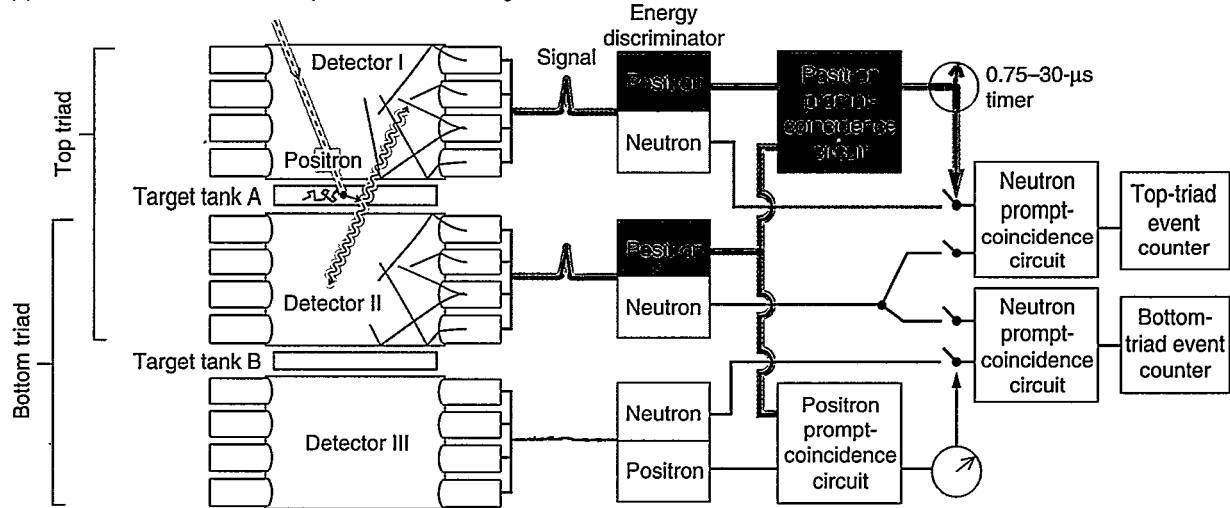
This flow diagram traces the generation of a set of delayed-coincidence signals in the top triad of the detector (target tank A and scintillation detectors I and II). An antineutrino (red dashed line) from the reactor has interacted with a proton in tank A through inverse beta decay, creating a positron and a neutron. As a result, two processes occur in tank A: positron annihilation, shown in diagram (a), and neutron capture, shown in diagram (b). In the case illustrated here, the delay between the two processes is 3 microseconds.

In diagram (a), the encounter between a positron and an electron in tank A results in two gamma rays, which go into scintillation detectors I and II, give up their energy, and produce a flash of visible light proportional to that energy. The photomultiplier tubes in each detector convert the light into an electronic signal, which is sent first to the positron signal discriminator and then to the positron prompt-coincidence circuit. The discriminator will accept the signals from detectors I and II if they are within the right energy range (between 0.2 and 0.6 MeV). The prompt-coincidence circuit will accept them if they arrive less than 0.2 microsecond apart. In this case, both conditions are fulfilled. The timer starts to tick and closes the switch to the neutron prompt-coincidence circuit for 30 microseconds, allowing signals from neutron capture to be recorded during that period.

The very large detector—over 2 meters high and weighing about 10 tons—had to be installed in those cramped underground quarters.

There was just enough room left for several preamplifiers (needed to boost the small signals from the photomultipliers) to be set on a rack near the detector, but the electronics had to be housed outside, in a trailer. The tank farm containing the precious liquid scintillator was also parked outside. The Los Alamos group used a whole network of stainless-steel pipes and valves, along with special pumps, to mix the solutions and pump them from the holding tanks in the parking lot into the detector down in the basement.

The team members stayed in Savannah River for over five months. They took data for about 900 hours when the reactor was on and for about 250 hours when it was off. Their immediate goal was to demonstrate a neutrino-like signal that was much larger when the reactor was on than off, indicating that it was caused by the flux of antineutrinos coming from the reactor.

In fact, the rate of delayed coincidences of the type described above was 5 times greater when the reactor was on than off and corresponded to about one reactor-associated event per hour. There was also the question of whether the delayed coincidences were accidental,

that is, caused by an accidental correlation between gamma rays and neutrons from the reactor. The neutron-capture delay time was unlikely to be more than 10 microseconds, whereas data were taken for up to 30 microseconds.

Thus, the accidental background rate could be estimated as the rate of delayed coincidences that occurred with neutron-capture delay times between 11 and 30 microseconds. Using this estimate, the team derived the rate of signal to accidental background events to have been 4 to 1.

Although the delayed-coincidence signal is a telltale signature of inverse beta decay, the Los Alamos team members took nothing for granted.

(b) T = 3 μs Neutron capture produces neutron signal.
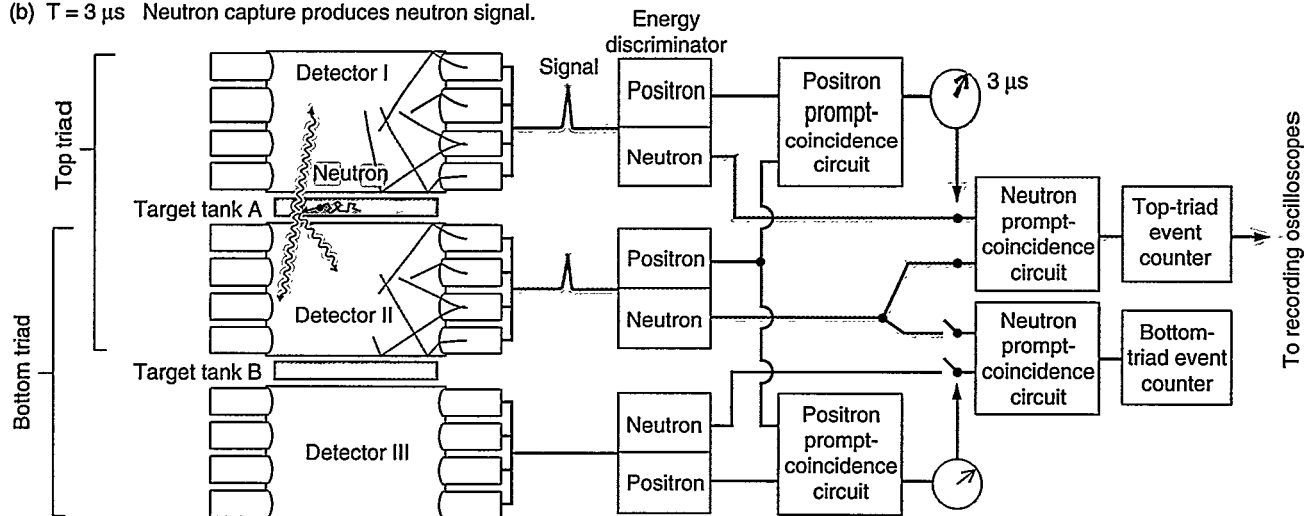
Diagram (b) pictures the slowdown of the neutron that had been generated simultaneously with the positron and its final capture by a cadmium nucleus in tank A. The excited cadmium nucleus drops to a lower energy state by emitting gamma rays, which once again create flashes of visible light in detectors I and II. The photomultiplier tubes detect that light and are shown to have produced two electronic signals whose energy is within the acceptable range, that is, the energy is greater than 0.2 MeV in each detector, with a total energy from 3 to 11 MeV (as determined by the discriminator). The signals are less than 0.2 microsecond apart in reaching the neutron prompt-coincidence circuit. Thus, they are accepted as a true signal of neutron capture. At this point, the timer has advanced to

3 microseconds, indicating the delay between the two processes. The delayed-coincidence signals caused by the neutrino-induced inverse beta decay is now complete. A scaler is automatically activated, the recording oscilloscopes are triggered to sweep across the cathode-ray screens, and the signals from the positron and neutron circuits, which have been stored on delay lines, are presented to the oscilloscopes.

They tested their measured signal extensively to ensure that it was indeed due to the products of neutrino-induced inverse beta decay, in particular that

• the first and second prompt-coincidence pulses were generated by positron annihilation and neutron capture, respectively, rather than other processes,

• the signal was proportional to the number of target protons, and

• the signal was not due to neutrons and gamma rays from the reactor.

For example, to check the positron signal, the Los Alamos researchers compared the pair of prompt-coincidence pulses making up the positron signal with those produced during a test run by a positron source (copper-64) dissolved in the water. To check the neutron capture signal, they doubled the amount of cadmium in the water to see if the average time delay between the positron-annihilation and neutron-capture signals decreased, as expected if the second signal was truly due to neutron capture.

To test that the signal was proportional to the number of target protons, they reduced the number of protons to half the original value by filling the tank with an equal mixture of heavy water (D_2O) and ordinary water. They then looked for a decrease in the signal corresponding to the decrease in the cross section for inverse beta decay on

deuterium versus the cross section for inverse beta decay on hydrogen.

Finally, to test whether the signal was coming from background neutrons and gamma rays from the reactor, they surrounded the detector with additional makeshift shielding. Bags of sawdust donated by a local sawmill and soaked in water for increased density were a cheap and flexible solution to the problem of creating an additional shield. Their effect was to decrease the reactor-associated accidental events, whereas the signal stayed constant. This and all other tests confirmed that the signal was indeed due to reactor antineutrinos being captured by protons in the water tanks of the detector and inducing inverse beta decays.
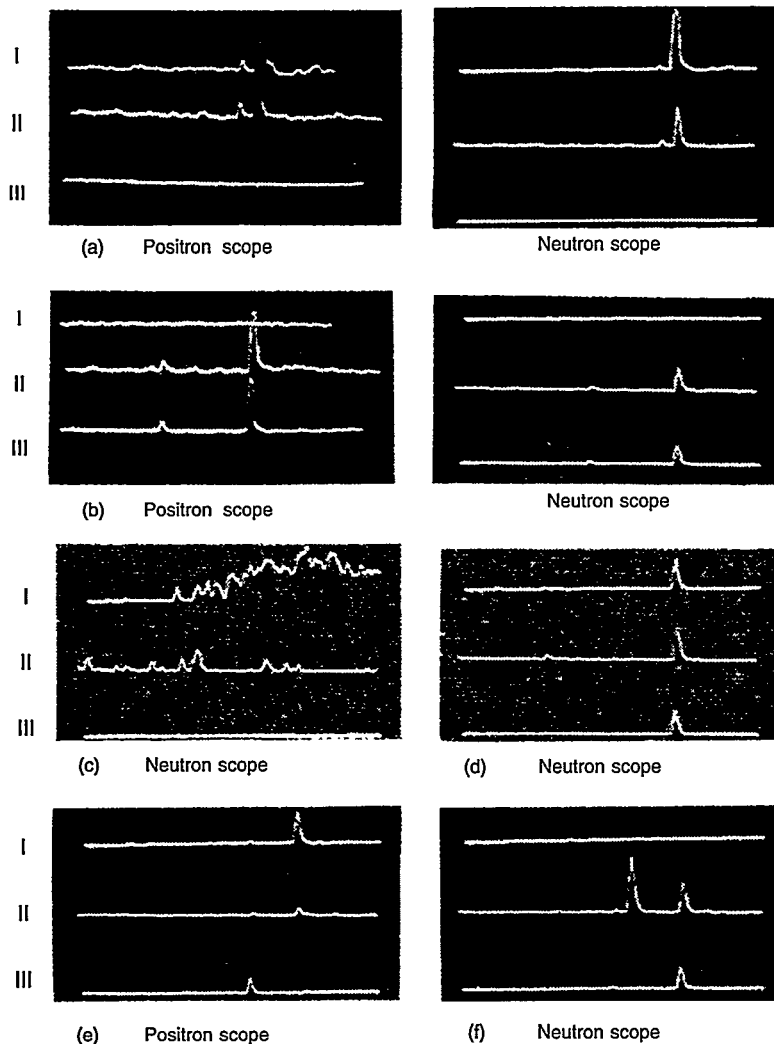
Figure 5. Oscilloscope Traces from the Savannah River Experiments
In these oscilloscope pictures, traces from detectors I, II, and III are labeled I, II, and III, respectively. The label under each frame indicates whether the signals were recorded by the scope for positron annihilation or the scope for neutron capture. Acceptable delayed-coincidence signals are shown in (a) and (b), while rejected signals are pictured in (c) through (f).

(a) The delayed-coincidence signal in these two frames has occurred in the top triad of the detector because the pulses appeared in detectors I and II. *Positron scope:* The pulse energies in detectors I and II were 0.30 MeV and 0.35 MeV, respectively. The pulses reached the positron circuit in prompt coincidence (less than 0.2 microsecond apart) and were accepted as a signal of positron annihilation. *Neutron scope:* The pulse energies in detectors I and II were 5.8 MeV and 3.3 MeV, respectively. These pulses arrived in prompt coincidence and were accepted as a signal of neutron capture. The delay between the positron and neutron signals was 2.5 microseconds. (b) The delayed-coincidence signal in these two frames has occurred in the bottom triad because the pulses appeared in detectors II and III. *Positron scope:* The pulse energies in detectors II and III were 0.25 MeV and 0.30 MeV, respectively.

Neutron scope: The pulse energies in detectors II and III were 2.0 MeV and 1.7 MeV, respectively. The delay between the positron and neutron signals was 13.5 microseconds. (c) The pulses from the neutron circuit were the result of electrical noise. (d) These three pulses from the neutron circuit were caused by a cosmic-ray event. (e) These three pulses from the positron circuit were caused by a cosmic-ray event. (f) These pulses may have been caused by a cosmic-ray event. They were rejected as a signal of neutron capture because of the extra pulse from detector II. Frames like this one occurred more often than would be expected from chance coincidences. They were, however, not often enough to affect the results considerably. These data appeared in Reines, Cowan, Harrison, et al. 1960.

Announcement of Results

On June 14, 1956, after all the tests had been completed, Reines and Cowan sent a telegram to Pauli at Zürich University:

"We are happy to inform you that we have definitely detected neutrinos from fission fragments by observing inverse beta decay of protons. Observed cross section agrees well with expected six times ten to minus forty-four square centimeters."

In his 1979 article in *Science* about the early days of experimental neutrino physics, Reines describes Pauli's reaction to the news:

"The message was forwarded to him [Pauli] at CERN, where he interrupted the meeting he was attending to read the telegram to the conferees and then made some impromptu remarks regard-ing the discovery. We learned later that Pauli and some friends consumed a case of champagne in celebration."

Although the intent of the Savannah River experiment was to get a positive signal of neutrino detection, the experiment also yielded a measurement of the rate, or more exactly the cross section, for inverse beta decay. (The cross section for the neutrino to be captured by a proton can be thought of as the effective

target area that a proton presents to a neutrino. The larger the area, the more likely it is that the process will occur.)

The measured rate, or the number of events per second, depends on (1) the rate at which neutrinos are entering the target area (the neutrino flux was approximately 1,013 neutrinos per square centimeter per second), (2) the number of target protons in the water tank (approximately 1,028 target protons), (3) the cross section for the reaction, and (4) the efficiency of the detectors in picking up positron and neutron signals from the reaction.

According to Fermi's theory, the cross section for inverse beta decay varies with energy. Given the energy spectrum of the reactor-produced antineutrinos (the average energy was 3 MeV), the theoretically predicted cross section for inverse beta decay on protons is 6.3×10^{-44}, with an uncertainty of about 25 percent arising from the uncertainty of the energy spectrum for the reactor neutrinos. The violation of parity conservation (namely, the symmetry between left-handedness and right-handedness) by the weak force had not yet been discovered, and so this theoretical value was based on the parity-conserving formulation of Fermi's theory of beta decay in which the neutrino, like the electron, has four independent degrees of freedom.

In July 1956, a brief article in *Science* by Reines, Cowan, Harrison, McGuire, and Kruse announced that the Savannah River experiment had confirmed the tentative findings of the Hanford experiment. The authors also stated that their results were in agreement within 5 percent of the theoretically predicted value for the inverse-beta-decay cross section. Such results were fortuitous given the uncertainties in the neutrino flux and in the detector efficiency.

A more detailed paper on this experiment published in *Physical Review* in 1960 reported a cross section twice as large as that reported in 1956. According to Reines (1979), the increase in the value occurred because "our initial

analysis grossly overestimated the detection efficiency with the result that the measured cross section was at first thought to be in good agreement with [the pre–parity violation] prediction."

The theoretical cross section had also doubled between 1956 and 1960 because of the discovery in 1957 of parity nonconservation in the weak interactions and the formulation of the two-component theory of the neutrino (see the box "Parity Nonconservation and the Massless Two-Component Neutrino" on page 32). So, the measured cross section reported in the literature remained in agreement with the theoretical prediction.

In addition, after the 1956 experiment, Reines and Cowan did another measurement with a new setup and, in a 1959 *Physical Review* paper, reported results for the cross section that were in agreement with the two-component neutrino, parity-nonconserving theory.

Over the years, there has been some skepticism about the differing published values. These feelings may have been responsible for the forty years that had passed before the discovery of the neutrino was recognized with the Nobel Prize. Nevertheless, the award is a clear recognition that the Savannah River experiment was an extraordinary accomplishment. Reines wished that Cowan had been alive to share the prestigious award with him. The elusive product of the weak force that can penetrate the earth and travel to the ends of the universe was finally observed stopping in its tracks. The neutrino became a tangible reality, and the experiment itself set a precedent for using the neutrino as an experimental tool.

Indeed, since the Reines-Cowan experiments, neutrino detection has produced some dramatic results. One was the 1963 experiment of Lederman, Schwartz, and Steinberger proving that a second (muon) neutrino was paired with the muon in the way the known (electron) neutrino was paired with the electron. That result not only earned the discoverers the Nobel Prize, but also

established the first hint of the second family of elementary particles (all three families are introduced in the primer, "The Oscillating Neutrino," on page 28).

Another was the detection of a burst of neutrinos from supernova 1987A (SN1987A)—twenty hits within 12 seconds in two enormous detectors located on opposite sides of the planet, both buried deep underground where one expects to see only one neutrino event per day. It was the unmistakable signature of an exploding star, and it provided extraordinary confirmation of the exotic notion that neutrinos, the most standoffish members of the pantheon of elementary particles, could drive the largest explosion ever witnessed by human beings.

And at present, neutrino data are accumulating from even more-modern neutrino detectors, some buried deep underground, some poised at accelerators, some awaiting completion, all dedicated to seeing whether the neutrinos, long purported to be massless particles, not only carry mass but also oscillate from one identity to another as they fly freely through space.

The world of physics owes much to Fred Reines for these developments. His single-minded dedication to the neutrino set an example, not only in the 1950s but throughout his career. And his courage to "think big" continued well after his tenure at Los Alamos. Reines was one of the critical cospokespersons for the construction of the huge IMB detector, a water-filled, 8,000-ton Cerenkov detector located in the Morton salt mine near Cleveland, Ohio. It was there that half of the events from SN1987A were detected and many of the data on the oscillation of atmospheric neutrinos were gathered.

Through this volume, Los Alamos National Laboratory takes pride in the accomplishments of Fred Reines, Clyde Cowan, Jr., and the teams of Laboratory workers who performed to the best of their ability in demonstrating the existence of the neutrino. And Fred Reines, in his gracious way,

has openly thanked the Laboratory:
"Looking back, we had much to
be thankful for. We had indeed been
in the right place at the right time.
The unlikely trail from bombs to
detection of the free neutrino could,
in my view, only have happened at
Los Alamos." (Reines 1982) ∎

Further Reading

Bethe, H. A., and R. F. Bacher. 1936. Disintegration and Nuclear Forces. *Review of Modern Physics* 8 (82): 184.

Cowan, C. L. 1964. Anatomy of an Experiment: An Account of the Discovery of the Neutrino. *Smithsonian Institution Annual Report:* 409.

Cowan, C. L., Jr., F. Reines, F. B. Harrison, E. C. Anderson, and F. N. Hayes. 1953. Large Liquid Scintillation Detectors. *Physical Review* 90 (3): 493.

Cowan, C. L., Jr., F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. 1956. Detection of the Free Neutrino: A Confirmation. *Science* 124 (3212): 103.

Crane, H. R. 1948. The Energy and Momentum Relations in the Beta-Decay, and the Search for the Neutrino. 1948. *Reviews of Modern Physics* 20 (1): 278.

Reines, F. 1979. The Early Days of Experimental Neutrino Physics. *Science* 203 (4375): 11.

———. 1982. Neutrinos to 1960—Personal Recollections. *Journal de Physique* 43 (supplement to 12): C8-237.

Reines, F., and C. L. Cowan, Jr., 1953. Detection of the Free Neutrino. *Physical Review Letters* 92: 330.

———. 1959. Free Antineutrino Absorption Cross Section. I. Measurement of the Free Antineutrino Absorption Cross Section by Protons. *Physical Review* 113 (1): 273.

Reines, F., C. L. Cowan, Jr., F. B. Harrison, A. D. McGuire, and H. W Kruse. 1960. Detection of the Free Antineutrino. *Physical Review* 117 (1): 159.

Reynolds, G. T., F. B. Harrison, and G. Salvini. 1950. Liquid Scintillation Counters. *Physical Review* 78 (4): 488.

Sutton, C. 1992. *Spaceship Neutrino.* New York: Cambridge University Press.

Frederick Reines is best known for his discovery of the nearly massless elementary particle, the neutrino. For this work, he was awarded the Nobel Prize in physics in 1995. Collaborating with Clyde Cowan, Jr., Reines determined conclusively the existence of the neutrino during experiments conducted at the Savannah River Plant in 1956. Subsequently, Reines devoted his career to investigating the properties and interactions of the neutrino as it relates not only to elementary particle physics but also to astrophysics.

This lifelong research produced a number of fundamental "firsts" credited to Reines. One of the most recent achievements, the codiscovery of neutrinos emitted from supernova 1987A (SN1987A), demonstrated the theorized role of the neutrino in stellar collapse. Reines captured the difficulty of this work vividly: "It's like listening for a gnat's whisper in a hurricane."

Significant other firsts include detecting neutrinos produced in the atmosphere, studying muons induced by neutrino interactions underground, observing the scattering of electron antineutrinos with electrons, detecting weak neutral-current interactions of electron antineutrinos with deuterons, and searching for neutrino oscillations (the possibility of neutrino transformation from one type to another). In addition, Reines and his coworkers have pursued for nearly forty years a program of experiments to test some of the fundamental conservation laws of nature, including conservation of lepton number (which would be violated in the decay of an electron or neutrino or in the change of lepton type) and conservation of baryon number, which would be manifested in the decay of the proton, as predicted by the Grand Unified Theories of elementary particles.

Reines was born in Paterson, New Jersey, on March 16, 1918. He earned his M.E. in mechanical engineering in 1939 and his M.A. in science in 1941 from Stevens Institute

---

Clyde L. Cowan, Jr., was born in Detroit, Michigan, on December 6, 1919. He earned his B.S. in chemical engineering at the Missouri School of Mines and Metallurgy (later to become part of the University of Missouri) in 1940 and his M.S. and Ph.D. in physics from Washington University in St. Louis, Missouri, in 1947 and 1949, respectively.

During the Second World War, Cowan joined the U.S. Army Chemical Warfare Service as a 2nd lieutenant and shortly thereafter left for England with the 51st Troop Carrier Wing. While he was stationed in England, Cowan was involved in making changes to the newly developed radar. For this significant work, he was later awarded the Bronze Star.

Soon after the war, Cowan returned to the United States where he was accepted as the first physics graduate student to Washington University. His thesis was an in-depth study of the absorption of gamma radiation. Soon after graduate school, Cowan realized that Los Alamos was the logical place for him to work, and in 1949 he joined the Laboratory as a staff member. Only two years later, Cowan became group leader of the Nuclear Weapons Test Division at Los Alamos.

In 1951, Cowan began a historic collaboration with Fred Reines. Its outcome was the successful detection of the neutrino during an experiment conducted at the Savannah River Plant in 1956. After this discovery, neutrino physics became seminal

of Technology in Hoboken, New Jersey. He received his Ph.D. in theoretical physics from New York University in 1944. That same year, he joined the Los Alamos Scientific Laboratory as a staff member, later to become group leader in the Theoretical Division, and was tasked to study the blast effects of nuclear weapons. In 1959, Reines became head of the Physics Department at Case Institute of Technology. At the same time, her served as consultant to Los Alamos and the Institute for Defense Analysis, as well as trustee of the Argonne National Laboratory. In 1966, however, Reines accepted a dual appointment as the first dean of physical sciences and physics professor at the University of California, Irvine. Four years later, he was appointed professor of radiological sciences at Irvine's Medical School. When Reines retired in 1988, he was Distinguished Emeritus Professor of Physics at Irvine.

For his outstanding work in elementary particle physics, Reines has received numerous honors and major awards. In 1957, he became fellow of the American Physical Society; in 1958, Guggenheim fellow; in 1959 Alfred P. Sloan fellow; in 1979, fellow of the American Association for the Advancement of Science; and in 1980, member of the National Academy of Sciences. In 1981, Reines received the J. Robert Oppenheimer Memorial Prize. He was presented the National Medal of Science by President Ronald Reagan in 1983, the Brunno Rossi Prize in 1989, the Michelson-Morley Award in 1990, the W. K. H. Panofsky Prize and the Franklyn Medal in 1992. He is a member of the American Academy of Arts and Sciences.

During a 1985 interview with *The New York Times*, Reines labored when he was asked to describe the significance of his discovery of the neutrino: "I don't say that the neutrino is going to be a practical thing, but it has been a time-honored pattern that science leads, and then technology comes along, and then, put together, these things make an enormous difference in how we live." And now, more than forty years after the discovery of the neutrino, Reines' scientific peers believe that this discovery made Reines a giant in his field.

to worldwide studies of the weak force. In 1957, Cowan was awarded a Guggenheim fellowship to study the physics of the neutrino and its interactions with atomic nuclei.

Cowan's creativity has been a mark of his scientific career from the early and fruitful years in Los Alamos to the successful teaching years at the Catholic University of America, where he was a physics professor from 1958 until his untimely death in 1974. Upon his suggestion, the bubble chamber became a tool for studying neutrino interactions. Cowan was one of the first physicists who used large scintillation counters for particle detection, an important technique in elementary particle physics. His collaboration with Reines led to the development of the whole-body counter, which measured low levels of naturally occurring radiation in humans. Having witnessed about thirty nuclear explosions while he was in the Nuclear Weapons Test Division at Los Alamos, Cowan was among the first to have studied the electromagnetic signal produced by a nuclear explosion.

Throughout his career, Cowan served as a consultant to the United States Naval Academy, the U.S. Atomic Energy Commission, the Naval Ordnance, and the Smithsonian Institution, where he helped create the permanent Hall of Nuclear Energy. Cowan was a fellow of the American Physical Society and the American Association for the Advancement of Science. He was a member of numerous scientific and civic organizations. Having dedicated his life to scientific investigation, Cowan has been a source of inspiration to generations of young, aspiring scientists.

The Oscillating Neutrino

An introduction to neutrino masses and mixings

Richard Slansky, Stuart Raby, Terry Goldman, and Gerry Garvey as told to Necia Grant Cooper
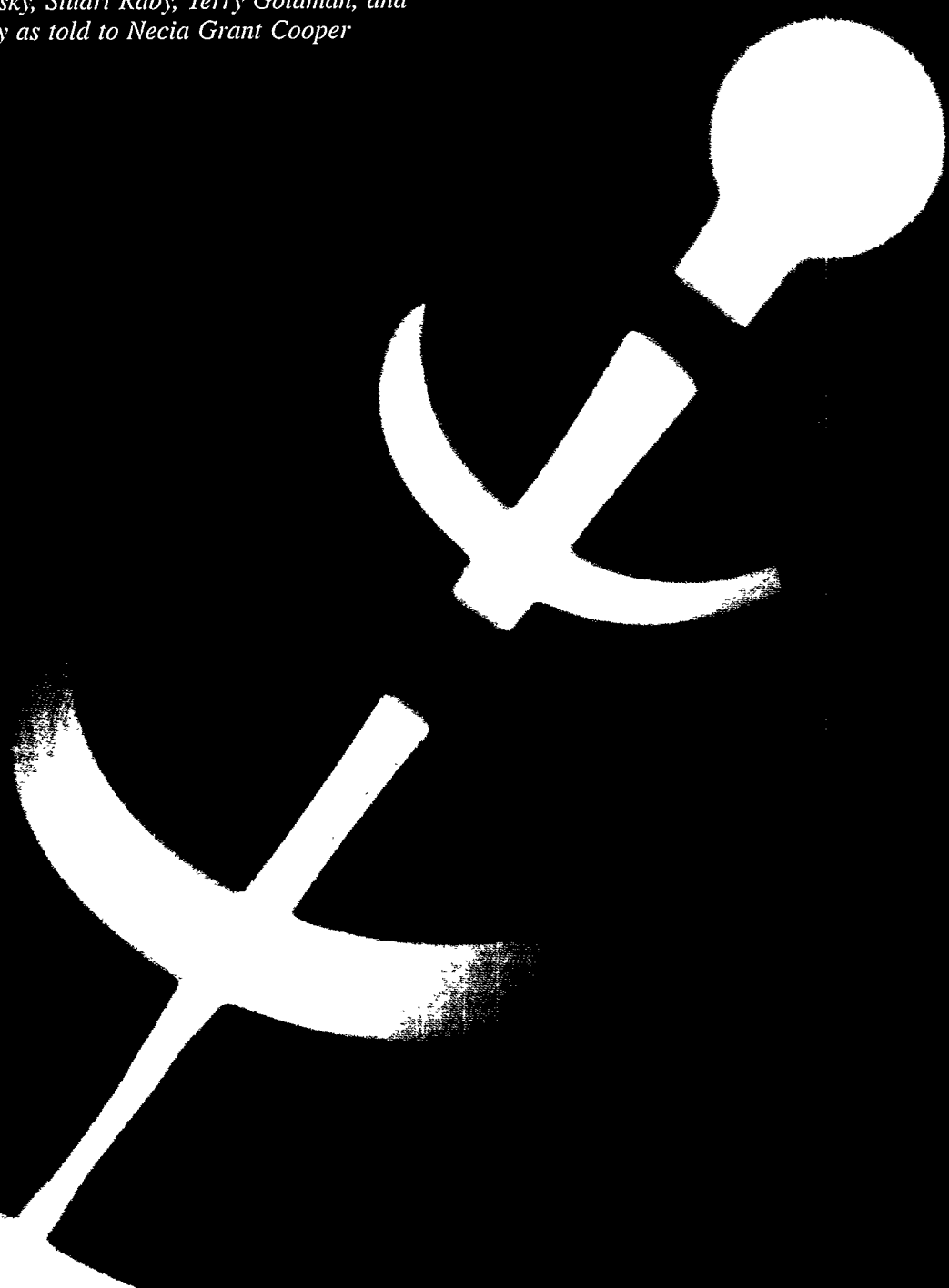
The creator of the neutrino is testing and teasing us. Moshe Gai

We do not know . . . [if] neutrinos are massive or massless. We do not know if the potentially massive neutrinos are Majorana or Dirac, and we do not know if these neutrinos can oscillate among flavours. . . In short, there is a great deal we do not know about neutrinos. Jeremy Bernstein, 1984.
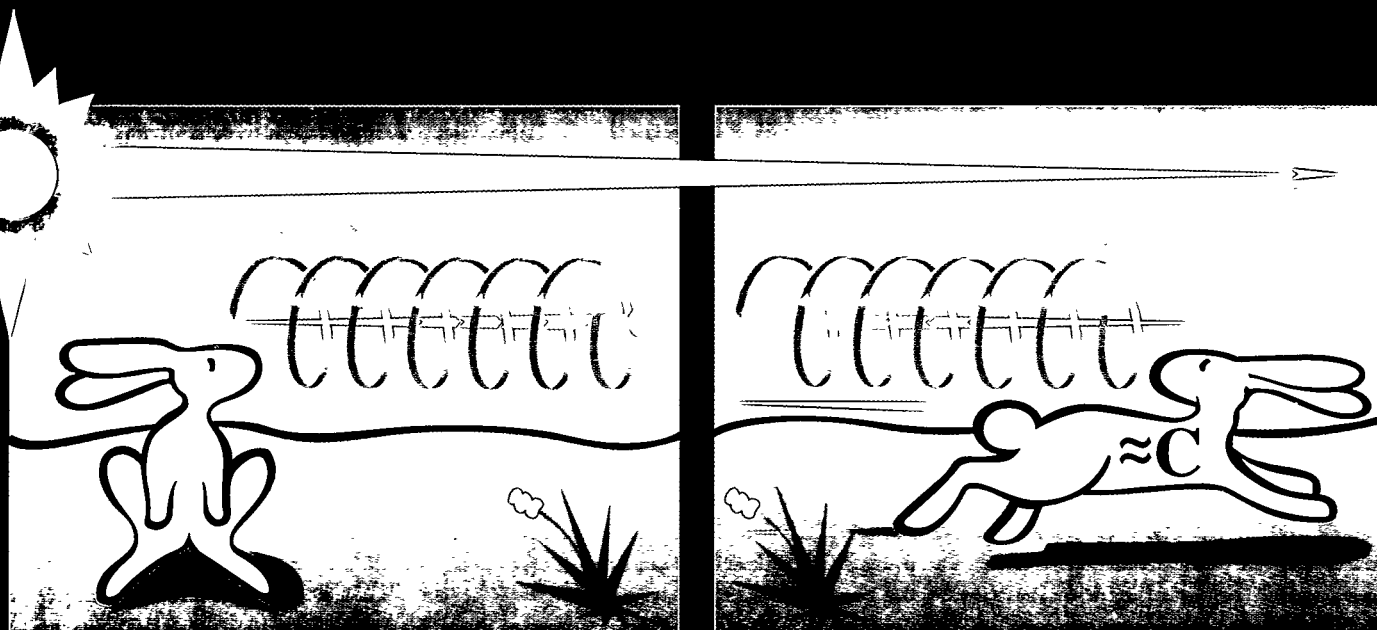
Looks left-handed.

No—right-handed?!*

The neutrino, the theoretical construct of sixty years ago, has acquired a presence in both physics and cosmology. It is both actor and probe. It explains numerous mysteries of the observable world. Yet every new characteristic it reveals opens up more questions about its true nature.

For decades, these bits of matter have been described as massless, left-handed particles: left-handed because they were always "spinning" counterclockwise in the manner of a left-handed corkscrew. But new evidence implies that neutrinos have very tiny masses and can spin in either direction. Remarkably, the new data also suggest that neutrinos might oscillate, or periodically present themselves as one of several different types.

The primer that follows explains why this strange behavior would fit in with theoretical expectations and how oscillations could reveal neutrino masses no matter how small. It also introduces questions that will become relevant. Why are neutrino masses so small? Do the very light neutrinos have very heavy relatives that make their masses small and give us hints of the new physics predicted by the Grand Unified Theories? Are neutrinos their own antiparticles? Do neutrinos have very light sterile relatives that provide a hiding place from all interactions? Physicists continue to chase after neutrinos, and every time these ghostly particles are caught, they seem to point toward new challenges and new possibilities.

*This neutrino must have mass.

After Reines and Cowan detected the neutrino in the late 1950s, particle physics went through a spectacular flowering that culminated in the formulation of the Standard Model. This model incorporates all that is known about the subatomic world. It identifies the most elementary constituents of matter, the elusive neutrino being among them, and then describes all the ways in which these elementary constituents can interact with and transform among each other. Awesome in scope, this theory provides a consistent picture of every realm of the physical world: from the hot, dense early universe resulting from the Big Bang to the thermonuclear furnace at the center of the Sun, from phenomena at the smallest, subatomic distance scales accessible at particle accelerators to those at the farthest reaches visible through the Hubble telescope. The same forces and symmetries and the same set of elementary building blocks seem sufficient to describe the underlying physics of all phenomena observed so far.

But for over two decades, ever since the Standard Model was initially formulated, expectations of "physics beyond the Standard Model" have been almost palpable among those familiar with the model's details: The theory just has far too many arbitrary parameters and mysterious relationships to be the final one. Now, after many years of searching, the first hard evidence for new physics may be at hand. The new physics—nonzero neutrino masses and "mixing" among the neutrinos from different families—has long been anticipated because it parallels the behavior seen among quarks. But still, it is quite exciting because it both affirms the central concepts of the Standard Model and appears to point toward the most popular extensions, the Grand Unified Theories.

This article introduces the neutrino in the context of the Standard Model and explains how the new data on neutrinos relate and suggest extensions to that theory.

Neutrinos in the Standard Model

The Standard Model identifies twelve building blocks of matter (see Figure 1), six quarks and six leptons (and their respective antiparticles). The quarks are the building blocks that have fractional electric charge and interact primarily through the strong nuclear force, also called the color force. Color binds quarks together to form the proton, the neutron, all nuclei, and all the other hadrons (strongly interacting particles). The charged leptons are the building blocks that interact through the other three forces of nature (weak, electromagnetic, and gravity) but never through the strong force. As a result, leptons are never bound inside the nucleus by the strong force. The leptons include the electron, the heavier "electron-like" muon and tau, and these three particles' neutral partners: the electron neutrino, the muon neutrino, and the tau neutrino. Among these twelve constituents, only the neutrinos are nearly or exactly massless.

Dubbed "the little neutral ones" because they have no electric charge, neutrinos interact with matter only through the weak force and gravity. Recall that the weak force creates neutrinos through beta decay (see the box "Beta Decay and the Missing Energy" on page 7). In that particular weak decay process, a neutron, either free or in a nucleus, transforms into a proton, and two leptons are created: an electron (or "beta" particle) and an electron antineutrino. More generally, the weak force is the force of transmutation, able to transform one type, or "flavor," of quark into another or one flavor of lepton into another. It is also the "weakest" known force (apart from gravity), about a hundred million times weaker than electromagnetism at "low" energies, which means that it acts a hundred million times more slowly. For example, unstable particles decay through the weak force in times on the order of 10^{-8} second, whereas the characteristic times for electromagnetic decays and

strong decays are 10^{-16} second and 10^{-23} second, respectively.

It is precisely this lack of interaction strength that makes the neutrino so elusive. For not only does the weak force create neutrinos, often through beta decay, but it also mediates the only processes that can absorb them.

The intimate connection between the weak force and the neutrino has sometimes made their separate properties difficult to sort out. In fact, the theory that the neutrino is massless and left-handed (and the antineutrino right-handed) was invented to explain why the weak force violates the symmetry known as parity, also called right-left, or mirror, symmetry. If the weak force conserved parity, any weak process and its mirror image would be equally likely. Instead, in 1956, C. S. Wu and coworkers observed a striking asymmetry in the beta decay of cobalt-60 (see the box "Parity Nonconservation and the Two-Component Neutrino" on page 32.) The asymmetry suggested that *all* the antineutrinos emitted in the decay had right helicity,[1] that is, they were "spinning" like right-handed corkscrews (rotating clockwise around their direction of motion). But in a universe with right-left symmetry, an equal number of antineutrinos should have been spinning counterclockwise, like left-handed corkscrews. The fact that only right helicity was observed is an example of "maximal" parity violation.

---

[1]Helicity is identical to handedness (or chirality) for massless neutrinos and nearly identical to handedness for particles traveling near the speed of light. For that reason, helicity is sometimes loosely referred to as "handedness." For massive particles, however, the two quantities are quite different. Massive particles must exist in right- and left-helicity and in right- and left-handed states. As illustrated in the box on page 32 and the cartoon on page 29, helicity is the projection of the spin along the direction of motion. It can be measured directly, but its value depends on the frame from which it is viewed. In contrast, handedness is a relativistically invariant quantity, but it is not a constant of the motion for a free particle and cannot be measured directly. Nevertheless, handedness is the quantity that describes the properties of the weak force and of the particle states that interact through the weak force and have definite weak charges.

First Family

Up u

Electric charge = +2/3.
Protons have two up quarks;
neutrons have one.

Mass ≈ 3 MeV/c^2.

Down d

Electric charge = −1/3.
Protons have one down quark;
neutrons have two.

Mass ≈ 6 MeV/c^2.

Electron e

Electric charge = −1.
Is responsible for electrical
and chemical reactions.

Mass = 0.511 MeV/c^2.

Electron Neutrino ν_e

Electric charge = 0.
Is paired with electrons by the
weak force. Billions fly
through us every second.
Mass = 0 (assumed).

Anti-up $\bar{u}$
Antidown $\bar{d}$

Positron e^+
Electron antineutrino $\bar{\nu}_e$

Second Family

Charm c

Electric charge = +2/3.
Is heavier than the u.

Mass ≈ 1,500 MeV/c^2.

Strange s

Electric charge = −1/3.
Is heavier than the d.

Mass ≈ 170 MeV/c^2.

Muon μ

Electric charge = −1.
Is heavier than the e.

Mass = 105 MeV/c^2.

Muon Neutrino ν_μ

Electric charge = 0.
Is paired with muons by the
weak force.

Mass = 0 (assumed).

Anticharm $\bar{c}$
Antistrange $\bar{s}$

Antimuon μ^+
Muon antineutrino $\bar{\nu}_\mu$

Third Family

Top t

Electric charge = +2/3.
Is heavier than the c.

Mass ≈ 175,000 MeV/c^2.

Bottom b

Electric charge = −1/3.
Is heavier than the s.

Mass ≈ 4,500 MeV/c^2.

Tau τ

Electric charge = −1.
Is heavier than the μ.

Mass = 1,782 MeV/c^2.

Tau Neutrino ν_τ

Electric charge = 0.
Not yet seen directly.
Assumed to be paired with
the tau by the weak force.
Mass = 0 (assumed).

Antitop $\bar{t}$
Antibottom $\bar{b}$

Antitau τ^+
Tau antineutrino $\bar{\nu}_\tau$

Quarks

Leptons

Antiparticles

Figure 1. Building Blocks of Matter in the Standard Model

The elementary building blocks of matter in the Standard Model are six quarks and six leptons, each carrying an intrinsic spin of 1/2. The first family contains one quark pair—the up and the down—and one lepton pair—the electron and the electron neutrino. These four particles make up the ordinary matter that is found on Earth and throughout most of the immediate universe. In particular, the proton is made of the quark triplet *duu*, and the neutron is made of the quark triplet *udd*. The second and third families are also composed of one quark pair and one lepton pair. Apart from the neutrinos, which are massless, the particles in the second and third families are more massive than their counterparts in the first family. They are also unstable and only stick around for tiny fractions of a second because the weak force allows them to decay into less massive particles. The more massive versions of the quarks and leptons are created in very high energy processes at the center of stars and galaxies, in high-energy accelerators, and at about 30 kilometers above the surface of the earth through the collision of very high energy cosmic rays (mostly protons) with molecules in the earth's atmosphere. The first hint that there are particles beyond the first family came in 1937 with the discovery of the muon. The top quark, the heaviest member of the third family, was not seen until 1995. And so far, the tau neutrino has not been detected directly. Nevertheless, the three families are so similar in structure that some of their members were anticipated long before they were observed. All particles have corresponding antiparticles (listed last in this figure) with opposite charge.

Parity Nonconservation and

(a) Four States of a Spin-1/2 Particle

Right-helicity particle

Left-helicity particle

Right-helicity antiparticle

Left-helicity antiparticle

Spin aligned with momentum

Spin opposite to momentum

p = Particle momentum

s = Spin pseudovector

(b) Mirror Reflection of a Right-Helicity Particle

Mirror image has left helicity

Particle has right helicity

The helicity of a particle relates its intrinsic spin to its direction of motion. All quarks and leptons, including the neutrino, carry 1/2 unit of intrinsic angular momentum s, or spin (measured in units of $\hbar$). Spin is quantized, and for a spin-1/2 particle, it has two values relative to any selected axis of quantization, which we choose to call the z-axis. The spin is often represented by a pseudovector (red arrow) that points up ($s_z = 1/2$) or down ($s_z = -1/2$) along the axis of quantization, depending on whether the particle is spinning clockwise or counterclockwise around that axis when it is viewed from below. Helicity uses the direction of motion, or the momentum p, as the axis of quantization, where helicity is defined as $\lambda = s \cdot p/|p| = \pm 1/2$.

As shown in (a), spin-1/2 particles usually have four independent states: the particle with right or left helicity and the antiparticle with right or left helicity. A particle has right helicity ($\lambda = 1/2$) if its spin and momentum point in the same direction. It has left helicity ($\lambda = -1/2$) if its spin and momentum point in opposite directions. The mirror image of a right-helicity particle is a left-helicity particle, as shown in (b). (Note that, being a pseudovector, s does not change direction under spatial inversions. Like total angular momentum J and orbital angular momentum l, it transforms as $r \times p$ does.) Until the 1950s, it was taken for granted that the laws of physics were invariant under a mirror reflection or an inversion of spatial coordinates (also called parity inversion). If parity were conserved, a spin-1/2 particle would exist in both left- and right-helicity states.

But in June of 1956, two young physicists, C. N. Yang and T. D. Lee, suggested that the weak force might violate parity conservation, and they outlined several types of experiments that could test their hypothesis. Six months later, C. S. Wu reported the results of one such experiment. Wu aligned the spins of cobalt-60 nuclei along an external magnetic field and measured the directions of the electrons emitted by those nuclei in beta decay:

$$^{60}\text{Co}_{27} \rightarrow {}^{60}\text{Ni}_{28} + e^- + \bar{\nu}_e \; .$$
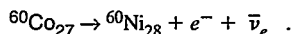
The electrons were almost always emitted in the direction opposite to the nuclear spins, as shown in (c). If parity were conserved, there should be no correlation between the spins and the momenta of the electrons emitted in the decay. A correlation between spin and

The results of the cobalt experiment were formalized in the theory of the two-component massless neutrino, according to which the antineutrino is always right-handed (or has right helicity), the neutrino is always left-handed (or has left helicity), and the neutrino is a massless particle (see the box above). But the weak force

itself was soon recognized to violate parity maximally because it acts on only the left-handed states of all quarks and leptons, whether they have mass or not. In other words, left-handedness is an intrinsic property of the weak force and not necessarily of the neutrino. Thus, in principle, the neutrino

could have a small mass.

Nonetheless, the original theory of the massless, left-handed neutrino was included in the "minimal" Standard Model, primarily because there was no evidence to the contrary. All direct measurements of neutrino masses have yielded only upper limits (see the article "Tritium Beta Decay

the Massless Two-Component Neutrino

momentum is measured by the average value of the dot product $s \cdot p$, which changes sign under a parity inversion and therefore must be zero if parity is conserved in a given process. The nearly perfect correlation of the nuclear spins and the electron momenta in the cobalt experiment was an example of *maximal* parity violation.

To explain the violation, Lee and Yang assumed that the antineutrino was always emitted with right helicity. As shown in (d), the decay decreases the nuclear spin by one unit. Aligning the spins of the electron and the antineutrino along the nuclear spin ($1/2 + 1/2 = 1$) will make up for this decrease. If the antineutrino always has right helicity (momentum and spin aligned), the electron will have to be emitted with left helicity (momentum opposite to spin) and in the direction opposite to the nuclear spin, which is just what was observed in the cobalt experiment. Yang and Lee formalized this interpretation in the theory of the two-component neutrino (1957), which postulates that the neutrino comes in only two forms, a left-helicity ($\lambda = -1/2$) particle and a right-helicity ($\lambda = 1/2$) antiparticle. But definite helicity has a profound consequence. To have left helicity in all coordinate systems moving with constant velocity relative to each other, as required by special relativity, the left-helicity neutrino must be traveling at the speed of light. Otherwise, one could imagine observing the particle from a coordinate system that is moving faster than the neutrino. As one zipped past, the neutrino's momentum would appear to be reversed, while its spin direction would remain unchanged. The neutrino would then appear to have right helicity! So, helicity remains independent of the reference frame only if the neutrino moves at the speed of light. But then the neutrino must be a massless particle. Helicity then becomes identical to the relativistically invariant quantity known as "handedness," so the neutrino is a *left-handed* massless particle.

The theory of the two-component massless neutrino fits nicely with the Gell-Mann and Feynman formulation (1958) of the left-handed weak force (also known as the $V-A$ theory, for vector current minus axial vector current, a form that violates parity maximally). In this theory, the weak force picks out the left-handed components of particles and the right-handed components of antiparticles. Since the neutrino interacts only through the weak force, the two missing components of the neutrino (the right-handed particle and the left-handed antiparticle) would never be "seen" and would be superfluous—*unless the neutrino had mass.*

(c) Maximum Parity Violation in the Cobalt-60 Experiment

Magnetic field

^{60}Co nuclei with spins polarized along magnetic field

Electrons emitted in lower half plane
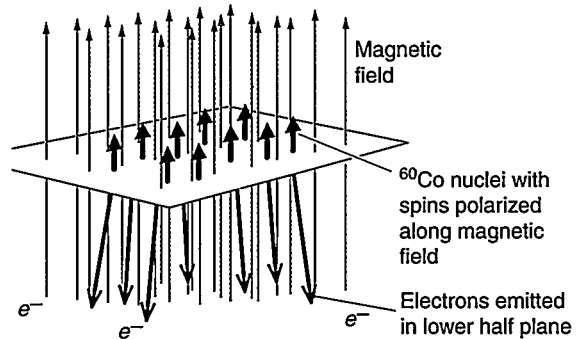
(d) Explanation of Cobalt-60 Experiment

$\bar{\nu}_e$ Right helicity

e^- Left helicity

$$J_z \quad 5 \quad = \quad 4 \quad + \quad \overbrace{1/2 + 1/2}$$

$$^{60}_{27}\text{Co} \longrightarrow {}^{60}_{28}\text{Ni} + e^-_L + \bar{\nu}_{eR}$$

and the Search for Neutrino Mass" on page 86). The assumption of massless neutrinos, however, has a consequence in the minimal Standard Model: It implies that lepton-family number is conserved. Each lepton family consists of a lepton pair. The electron and its neutrino constitute the electron family; the muon and its neutrino the muon

family; and the tau and its neutrino the tau family (refer again to Figure 1). Each lepton family is part of a much larger family that also includes the quarks and the respective antiparticles of the leptons and quarks.

Conserving lepton-family number means preserving strict boundaries between the electron, muon, and tau

families. For example, the muon and the muon neutrino can transmute into each other through the weak force (no change in the muon-family number), but the muon cannot decay directly into an electron. Instead, a member of the muon family (the muon neutrino) must also be produced during muon decay. Similarly, a tau cannot decay directly

Figure 2. The Electromagnetic Force

The electromagnetic force is transmitted through the exchange of the photon, the gauge particle for the electromagnetic field.

(a) Electron scattering. Two electrons (straight arrows) interact with each other through the exchange of a virtual photon (wiggly line). The direction of time is from left to right. The dots represent interaction vertices, where one electron emits a photon and the other electron absorbs it.

Initial state   Final state

Electron e^- e^- Electron

γ Photon

Electron e^- e^- Electron
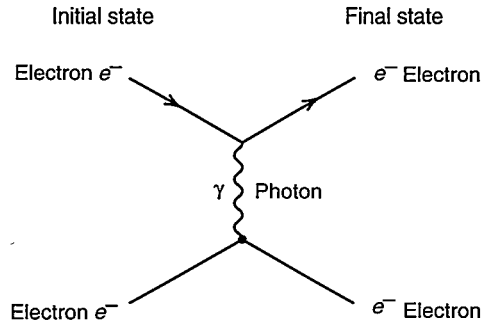
(b) Electron-positron annihilation and creation. When the diagram above is rotated by 90°, it represents an electron and positron that annihilate at the left interaction vertex to become a virtual photon, and then the virtual photon converts back to an electron and positron at the right interaction vertex. Note that, when an arrow points backward in time, it represents the antiparticle moving forward in time.

Initial state   Final state

Electron e^- e^+ Positron

Photon
γ

Positron e^+ e^- Electron
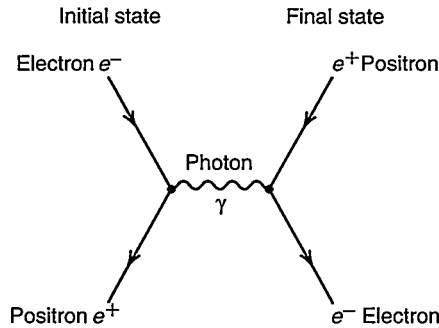
(c) The interaction vertex. All processes involving the electromagnetic force can be built up from the basic interaction vertex. In the left vertex, an electron emits or absorbs a photon, and in the right vertex, a photon turns into an electron-positron pair or vice versa.

Electron e^- e^- Electron e^+ Positron

Electron emits
or absorbs a photon. γ Photon

Photon
γ

Photon turns into an
electron-positron pair,
or the pair annihilates
into a photon. e^- Electron
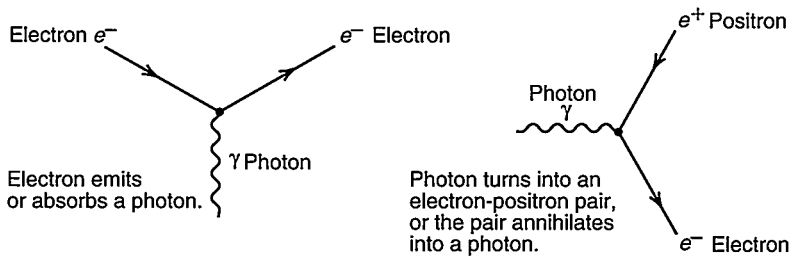
into a muon or an electron unless a tau neutrino is also produced. Finally, conservation of lepton-family number means that an electron neutrino cannot change into a neutrino from another family, or vice versa. These predictions of the minimal Standard Model have held up to increasingly precise tests.

Recent evidence, however, is changing this picture. Data from the neutrino oscillation experiment at Los Alamos, known as LSND (for liquid scintillator neutrino detector), as well as from

solar- and atmospheric-neutrino experiments, suggest that muon neutrinos can periodically change into electron neutrinos, and vice versa, as they travel through the Sun or even through empty space. One consequence would be that electron neutrinos arriving at Earth from the center of the Sun would appear to be too few in number although, in fact, the right total number would be present. Some would be "invisible" because they would have temporarily changed into another flavor—into muon or tau neutri-

nos whose interactions are unobservable in the detectors being used. As shown in later sections, this oscillation from one flavor to another can happen only if the different neutrino types have different masses, so measurement of oscillation is proof that neutrinos have mass.

Positive results from the LSND oscillation experiment have therefore caused a stir in the physics community. These results could explain the "solar-neutrino puzzle" (the apparent deficit in the number of solar neutrinos) and could also have impact on other topics in astrophysics and cosmology that involve large numbers of neutrinos. On a more abstract note, nonzero neutrino masses and oscillations among flavors would parallel the properties and behaviors seen among the quarks and would thus point toward a greater symmetry between quarks and leptons than now exists in the Standard Model. They might even point toward a more encompassing and unifying symmetry that has been anticipated in the Grand Unified Theories, in which quarks and leptons are different aspects of the same field and the strong, weak, and electromagnetic forces are due to a single symmetry.

Gauge Symmetries in the Standard Model

The Standard Model is built almost entirely from symmetry principles, and those principles have enormous predictive power. Symmetry means an invariance of the laws of physics under some group of transformations. And in the formalism of quantum field theory, the invariance implies the existence of a conserved quantity. One example is the group of rotations. We take for granted, and know from high-precision measurements, that space has no preferred direction and that we can rotate an isolated system (a group of atoms, a solar system, a galaxy) about any axis (or rotate the coordinates we use to describe that system) and not change the laws of physics observed by that system. This property is called

rotational invariance, and it has the profound consequence that the total angular momentum of an isolated system is conserved and therefore never changes. Similarly, if a system is invariant under time translations, its total energy is conserved. If a system is invariant under spatial translations, its total linear momentum is conserved.

In addition to these space and time symmetries, the Standard Model has certain powerful internal symmetries, called local gauge symmetries, that define both the charges of the quarks and leptons and the specific nature of the forces between them. Just as cubic symmetry implies the existence of four corners, six faces, and a group of rotations that interchange the position of the cube's faces and corners, the internal symmetries of the Standard Model imply that (1) the quarks and leptons fall into certain groups or particle multiplets, (2) the charges of the particles in each multiplet are related in a definite way, and (3) there is a group of internal rotations that transform one member of each multiplet into other members of that same multiplet.

But there is much more. Local gauge symmetries are those in which the magnitude of the transformation can vary in space and time. If the results of experiments are to stay invariant under such transformations (which is what symmetry means), gauge particles must exist that transmit or mediate the forces between the quarks and leptons. One quark or lepton emits a gauge particle, and another quark or lepton absorbs it. Through this exchange, each "feels" the force of the other. Further, the interac-

tion between the gauge particle and the quark or lepton actually causes one of the internal rotations defined by the local gauge symmetry, that is, the emission or absorption of a gauge particle causes that quark or lepton to transform into another member of the same multiplet. To give a geometrical analogy, if the world were perfectly symmetrical and the quarks and leptons in Figure 1 were like the faces of a cube, then the action of the gauge particles would be to rotate, or transform, one face (quark or lepton) into another.

The Electromagnetic Force. In the Standard Model, each force (strong, weak, and electromagnetic) is associated with its own local gauge symmetry, which, in turn, defines a set of charges and a set of gauge bosons that are the "carriers" or mediators of the force between the charged particles. The electromagnetic force is the simplest to describe. Figure 2(a) shows that two electrons, or any particles carrying electric charge, interact by the exchange of a photon, the gauge boson for the electromagnetic field. The exchange process can be pictured as a game of catch: One electron emits (throws) a photon, the other electron absorbs (catches) it, and the net result is that the two particles repel, or scatter from, each other. The classical electromagnetic field that explains how particles can interact at a distance is thus replaced by the exchange of a gauge particle.[2] The photon, of course, exists as an independent particle and can itself transform into a particle-antiparticle pair, most often an electron-positron pair. Figure 2(b) illustrates this process, and Figure 2(c) shows the basic interaction vertex. The local symmetry implies that all possible interactions involving the photon and electrically charged particles can be built up from this basic interaction vertex.

Finally, the local gauge symmetry holds only if the photon is identically massless, and the symmetry guarantees that electric charge is a conserved quantity, that is, the sum of the electric

charges before a reaction equals the sum of the charges after the reaction.

The Strong Force. The local gauge symmetry for the strong force is called color symmetry. The color charge has three distinct aspects that, for convenience, are labeled red, green, and blue (no relation to real colors is intended). The gauge particles are called gluons, the quarks can carry any of the three color charges, and two colored quarks interact and change color through the exchange of one of the eight colored gluons. Like the gauge symmetry for the electromagnetic force, the gauge symmetry for the strong force implies that the gluons are massless and that the total color charge is conserved. Because the gluons carry color, and are thus like electrically charged photons, the strong force is highly nonlinear and has some very bizarre properties. One is that quarks can never appear individually, and another is that all observable states of quarks and antiquarks (protons, neutrons, pions, and so forth) are colorless bound states, that is, they have no net color charge.

In the discussions that follow, we can ignore the strong force because leptons do not carry the color charge and the part of the Standard Model that describes the color interactions of the quarks (known as quantum chromodynamics) will not be affected by new data on neutrino masses and mixings.

The Weak Force. The Standard Model identifies two local gauge symmetries for the weak force and, therefore, two types of weak charges (weak isotopic charge and weak hypercharge). As a consequence, there are two types of gauge particles, the W and the Z^0 bosons, that carry the weak force between particles with weak charges. The neutrino, although electrically neutral, carries both weak isotopic charge and weak hypercharge and thus interacts with matter through the exchange of either the W or the Z^0.

Let us first consider the processes mediated by the W. This gauge boson

Figure 3. Beta Decay and Other Processes Mediated by the W

The W is the charged gauge particle of the weak force, so processes mediated by the W involve the exchange of one unit of electric charge. Quarks and leptons therefore change their identities through the emission or absorption of the W. In all the processes shown here, the arrow of time is from left to right, and an arrow pointing backward represents an antiparticle moving forward in time. The arrow on the W indicates the flow of electric charge. Note also that in each of these processes, electric charge is conserved at every step.

$$n \rightarrow p + e^- + \bar{\nu}_e$$

$$\mu^- \rightarrow \nu_\mu + e^- + \bar{\nu}_e$$
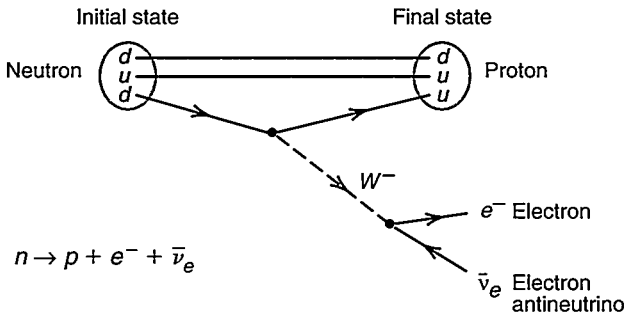
(a) Neutron beta decay. A neutron decays to a proton when a d quark in the neutron emits a W^- and transmutes into a u quark. Like the photon, the W^- can decay into a particle and an antiparticle, but here the particle is the electron, and the antiparticle is the electron antineutrino.
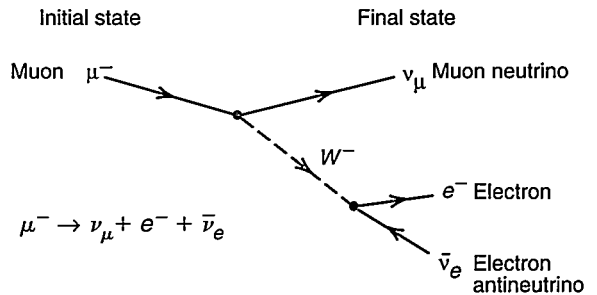
(b) Muon beta decay. This process is exactly analogous to the beta decay of the neutron. The muon transforms into a muon neutrino as it emits a W^-; the W^- decays into an electron and an electron antineutrino.

$$\bar{\nu}_e + p \rightarrow n + e^+$$

$$e^- + p \rightarrow n + \nu_e$$
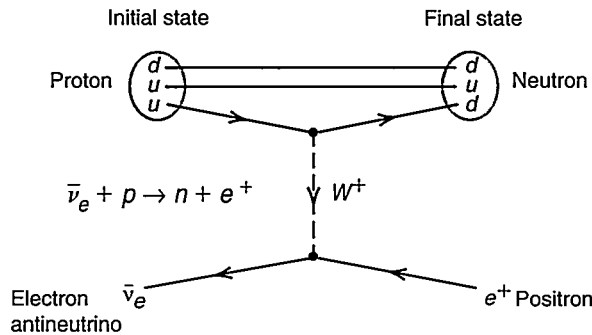
(c) Inverse beta decay. An electron antineutrino interacts with a proton by exchanging a W^+. The u quark emits a W^+ and transmutes to a d quark (thus the proton turns into a neutron). The electron antineutrino transmutes into a positron as it absorbs the W^+.
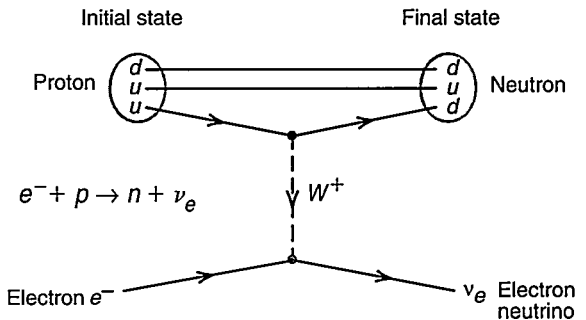
(d) Electron capture. This process is similar to inverse beta decay, except that an electron interacts with the proton. The electron transmutes into an electron neutrino as it absorbs the W^+.
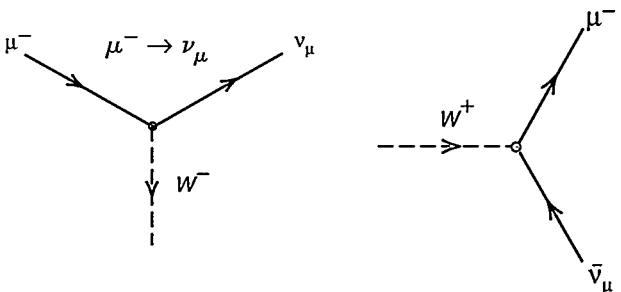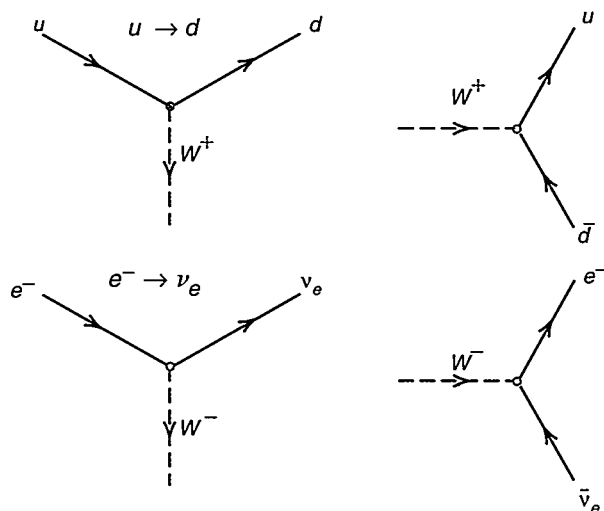
(e) Basic interaction vertices of the charged-current weak force. All the processes illustrated above can be built from variations of the interaction vertices shown here. These are analogous to the vertices shown in Figure 2 for the electromagnetic force, except here the gauge particle, the W, carries one unit of electric charge.

comes in two forms, the W^+ and the W^-. Each carries one unit of electric charge (plus or minus, respectively), so that when a particle carrying the weak isotopic charge emits or absorbs a W, it gains or loses one unit of electric (and weak isotopic) charge. The particle thereby changes its identity. Figure 3 illustrates neutron beta decay, muon beta decay, inverse beta decay, and electron or positron capture, all of which are processes mediated by the W.

The transmutation of the down quark into the up quark through emission of the W^- is the origin of the transmutation of a neutron into a proton in ordinary beta decay. In inverse beta decay, the process used by Reines and Cowan to detect the electron antineutrino, an up quark transmutes into a down quark as it emits a W^+, and an electron antineutrino transmutes into a positron as it absorbs that W^+. Because of the exchange of electric charge, the processes involving the exchange of the W are called "charged-current" weak processes. They are to be contrasted with the "neutral-current" processes mediated by the Z^0, in which no electric charge is exchanged. Note that this picture of the weak force, in which particles interact at a distance through the exchange of the W, modifies Fermi's original current-current theory of beta decay, in which two currents interacted at a point (see the box "Fermi's Theory of Beta Decay and Neutrino Processes" on page 8). The distance over which the W is exchanged is very short, on the order of 10^{-16} centimeter, which is substantially less than the diameter of a proton.

The scattering of electron neutrinos by electrons is a purely leptonic reaction that illustrates both charged-current and neutral-current modes (see Figure 4). In charged-current scattering, the electron emits a W^- and loses one unit of negative electric charge to become an electron neutrino. At the other end of this exchange, the electron neutrino absorbs the W^- and gains one unit of negative charge to become an electron. The initial and final particles are the same, but each has been transmuted into the other

Figure 4. Electron Neutrino–Electron Scattering

The electron and the electron neutrino can interact through the exchange of either the W, shown in (a), or the Z^0, shown in (b).
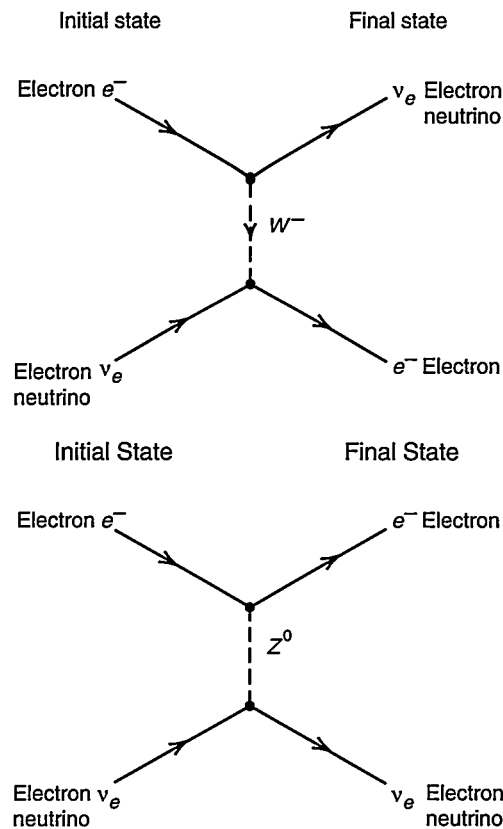
(a) Charged-current scattering. An electron and an electron neutrino interact through exchange of a W^-. The electron transmutes into an electron neutrino as it emits a W^-, and the electron neutrino transmutes into an electron as it absorbs a W^-.

(b) Neutral-current scattering. The electron and electron neutrino interact through exchange of the Z^0. No charge is transferred, and the particles maintain their identities.

through the charged-current weak interaction. In neutral-current scattering, the electron neutrino emits the Z^0, and the electron absorbs it. The two particles scatter from each other, but each maintains its identity as in electromagnetic scattering. All neutrino types can interact with electrons through neutral-current scattering, but only electron neutrinos can interact with electrons through charged-current scattering. That additional interaction may be important in enhancing the oscillation of electron neutrinos that exit the Sun (see the article "MSW" on page 156).

It is not coincidental that neutral-current scattering resembles electro-magnetic scattering. One of the great successes of the Standard Model was to show that the weak force and the electromagnetic force are related. The two types of weak charges, when added together in a specific linear combination, are equal to the electric charge.

Consequently, most quarks and leptons carry both types of weak charge as well as electric charge and can interact through exchange of the photon, the W, or the Z^0. For the neutrino, however, the specific sum that equals the electric charge (and couples to the photon) is zero, so that the neutrino is electrically neutral. (The electroweak theory, which describes the electromagnetic and weak forces, unifies the description of the photon and the Z^0 in a complicated way that will not be discussed in this article.)

Now, let us consider the particle multiplets that are consistent with the local symmetries of the weak force. Figure 5 lists the quarks and the leptons, along with their weak and electric charges. These particles fall naturally into three families (columns) consisting of a pair of quarks and a pair of leptons. Each pair is a doublet whose members transform into each other under the rotations of the local symme-

(a) Weak Doublets

First Family

Q I_3^w

$+\frac{2}{3}$ $+\frac{1}{2}$ u'

$-\frac{1}{3}$ $-\frac{1}{2}$ d'

Quarks

0 $+\frac{1}{2}$ ν_e

-1 $-\frac{1}{2}$ e

Leptons

Q = Electric charge

Second Family

Q I_3^w

$+\frac{2}{3}$ $+\frac{1}{2}$ c'

$-\frac{1}{3}$ $-\frac{1}{2}$ s'

Quarks

0 $+\frac{1}{2}$ ν_μ

-1 $-\frac{1}{2}$ μ

Leptons

I_3^w = Weak isotopic charge

Third Family

Q I_3^w

$+\frac{2}{3}$ $+\frac{1}{2}$ t'

$-\frac{1}{3}$ $-\frac{1}{2}$ b'

0 $+\frac{1}{2}$ ν_τ

-1 $-\frac{1}{2}$ τ

= Transmutation by the W

(b) Charged-Current Weak-Interaction Vertices

First Family

u' d'

W^+

ν_e e^-

W^+

Second Family

c' s'

W^+

ν_μ μ^-

W^+

Third Family

t' b'

W^+

ν_τ τ^-

W^+
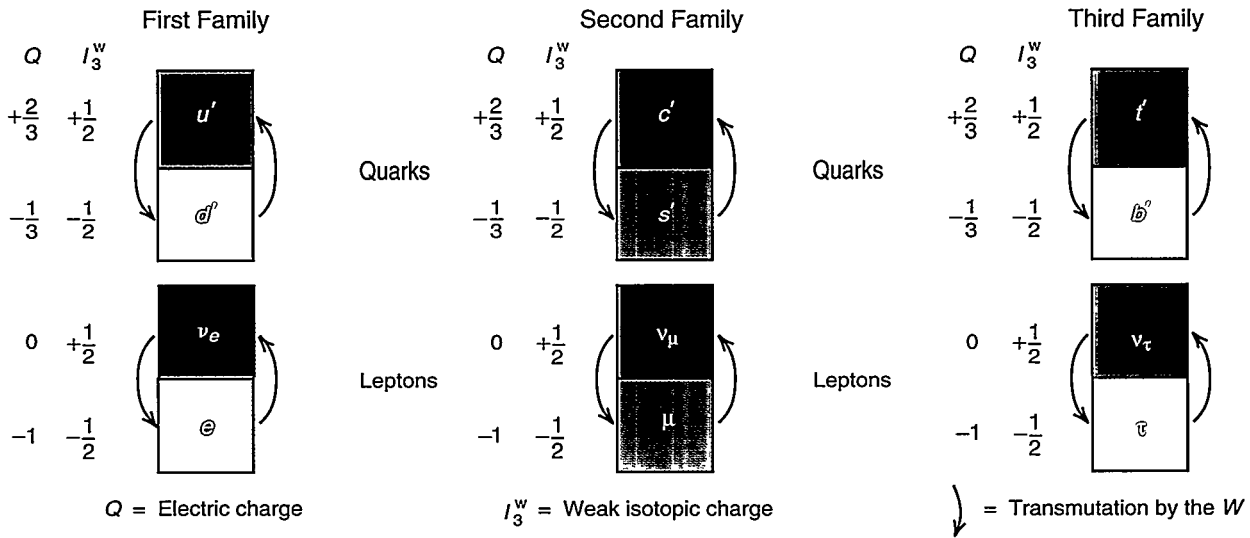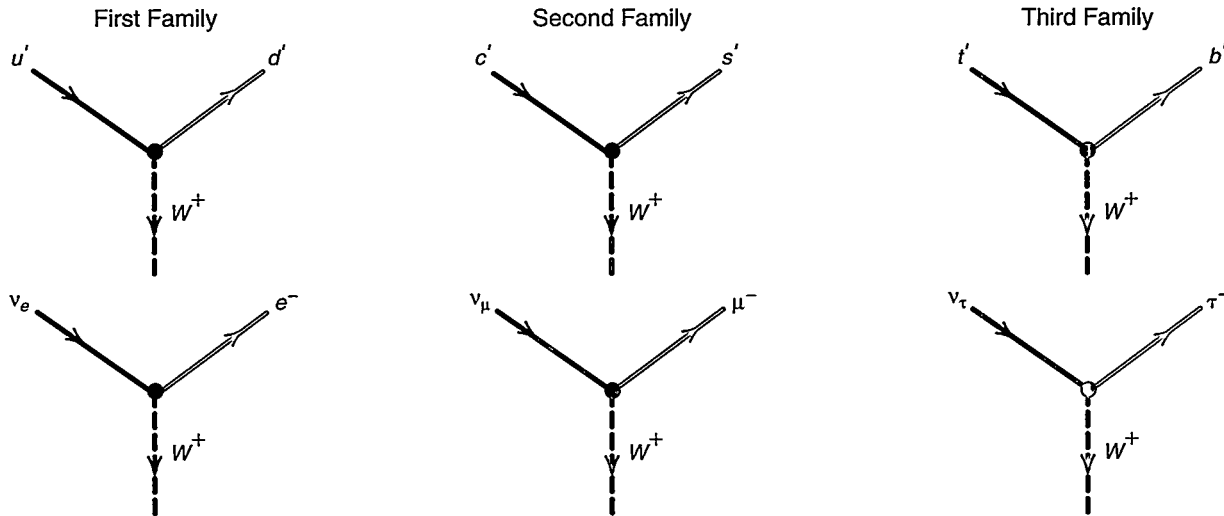
Figure 5. The "Weak" States—Particles Defined by the Local Symmetries of the Weak Force

The particle doublets, defined by the weak isospin symmetry of the weak force, are listed in (a) along with their electric charge Q and weak isotopic charge I_3^w. The particles fall into three families, each containing a quark weak isospin doublet and a lepton weak isospin doublet. (For each weak particle doublet, there is also a corresponding weak antiparticle doublet that is not shown.) The quarks in each doublet have been labeled with primes, u' and d' for example, to indicate that the weak quark states are distinct from the quark states shown in Figure 1. (The distinction will be made clear in the text.) As indicated by the red arrows and also expanded in (b), one member of the

doublet transforms into the other member by absorbing or emitting the W, the gauge particle for the charged-current weak interaction. Only the left-handed particles (or right-handed antiparticles) carry the weak isotopic charge and are members of the doublet. In the quark sector, the u' quark transforms into the d' quark and vice versa, the c' quark transforms into the s' quark and vice versa, and the t' quark transforms into the b' quark and vice versa. In the lepton sector, the electron and the electron neutrino transform into each other through interaction with the W, as do the muon and muon neutrino and the tau and tau neutrino. This universal interaction with

the W means that the muon and the muon neutrino or the tau and the tau neutrino could replace the electron and its neutrino wherever the latter pair appears in the charge-changing weak processes in Figures 3 and 4. (Whether those processes actually occur with the heavier leptons depends on the available energy.) The similarities among the weak isospin pairs extend to the electric-charge assignments as well. In each quark doublet, one member has electric charge $+2/3$; the other, $-1/3$. In each lepton doublet, one member has charge zero, and the other has charge -1. The weak-isotopic-charge assignments are likewise maintained from family to family.

try group (called weak isospin)[3]. The red arrows indicate the transmutation of one member into the other through absorption or emission of the W. Thus, members of a weak isospin doublet are like the two faces of a coin, and interaction with the W flips the coin from one face to the other. (Note that, for each weak particle doublet, there is a corresponding weak antiparticle doublet. Because the charge-changing weak force is left-handed, the particle doublets include only the left-handed components of the particles, whereas the antiparticle doublets include only the right-handed components of the antiparticles. That technicality becomes important in the later discussion of mass.)

Alternatively, if the W is emitted and not absorbed by another weakly charged particle, it decays into one member of the doublet and the *antiparticle* of the other member of the doublet. This occurs, for example, in beta decay (refer to Figure 3). The W^- is not absorbed but decays to an electron and an electron *antineutrino* ($W^- \to e^- + \bar{\nu}_e$). Likewise, the W^+ can decay to a positron (antielectron) and an electron neutrino ($W^+ \to e^+ + \nu_e$).

This brief introduction to the forces associated with and derived from the local gauge symmetries needs one crucial addition. The local gauge symmetries of the weak force are *not* exact symmetries of nature, and one sign of the symmetry breaking is that, unlike the photon and the gluons, which must be massless to preserve the local gauge symmetry, the W and the Z are very massive, weighing about 100 times the mass of the proton. More precisely, the mechanism that gives mass to the particles breaks the weak symmetries and, in certain situations, causes the weak charges not to be conserved. However, processes mediated by the weak gauge particles do conserve the weak charges, and the original symme-

---

[3]Weak isospin symmetry is an example of the special unitary symmetry in two dimensions, SU(2). Rotational symmetry, which leads to the allowed states of angular momentum, is another example of SU(2) symmetry.

try is apparent in them. Also, the local gauge symmetry specific to electromagnetic interactions is not affected at all by the symmetry breaking, and the electric charge is always conserved.

These complications notwithstanding, each of the forces in the Standard Model is derived from a local gauge symmetry; the three forces, therefore, look similar in that they act through the exchange of gauge bosons. The reliance on local gauge symmetries has worked so well that many theorists have tried to extend this idea even further. They are trying to find a local gauge symmetry that combines into one all the separate symmetries associated with the strong and electroweak forces. Such theories predict that the quarks and leptons within a family will fall into one multiplet, one set of particles that transform into each other through the gauge particles associated with the local symmetry. This effort is called grand unification and is the basis of the Grand Unified Theories.

It is remarkable that theorists anticipated not only the structure of the electroweak force, but also the existence of the charmed quark (the partner to the strange quark) and later the top and the bottom quarks by identifying the correct local gauge symmetry for the weak force and then predicting that all quarks and leptons form doublets under that symmetry.

The symmetry of the weak force was not immediately apparent from experiment for several reasons. For example, the charm, top, and bottom quarks and the tau particle are very heavy. They were not observed at low energies, and thus half of all family members were not known to exist. Also, the physical quarks listed in Figure 1 are not identical with the members of the quark doublets listed in Figure 5. In the next section, we examine what is known about the differences between those two sets of quark states because there is a strong possibility that leptons may be described by two sets of states analogous to the quark sets. In that case, neutrino oscillations would be predicted. These analogous properties of quarks

and leptons are expected in all theories in which these particles are relatives and can transmute into each other.

The Mysteries of Masses and Families in the Standard Model

Figures 1 and 5 reflect two different ways of defining and placing particles in families: The families in Figure 1 contain particles with definite mass (the unprimed quarks), whereas those in Figure 5 contain particles defined by the local gauge symmetries (the primed quarks). These local gauge symmetries provide the guiding principles in the construction of the Standard Model and in most extensions to it; therefore, the weak states described in Figure 5 offer a fundamental starting point in shaping our understanding of the fundamental particles.

If we were to ignore the masses of the particles and focus on the symmetries, each family would look like a carbon copy of the other two. In other words, each particle would have a "clone" in each of the other two families that has identical weak and electric charges and that has a partner with which it forms a doublet under the weak force. The quark clones are u', c', and t' and their weak partners d', s', and b', respectively. The lepton clones are e, μ, and τ and their weak partners ν_e, ν_μ, and ν_τ, respectively.

In fact, if the local gauge symmetries of the weak force were exact, the quarks and leptons would all be massless. There is no way to include in the theory a "mass" term that remains invariant under those local symmetries. (The general features of mass terms for spin-1/2 particles are described in the sidebar "Neutrino Masses" on page 64.) In reality, particles do have mass, and thus the Standard Model contains a symmetry-violating mechanism known as the Higgs mechanism. This mechanism was specifically introduced into the Standard Model to explain the masses of the weak gauge particles,

Families of Mass States

Families of Weak States

Figure 6. A Comparison of Mass States and Weak States in the Standard Model
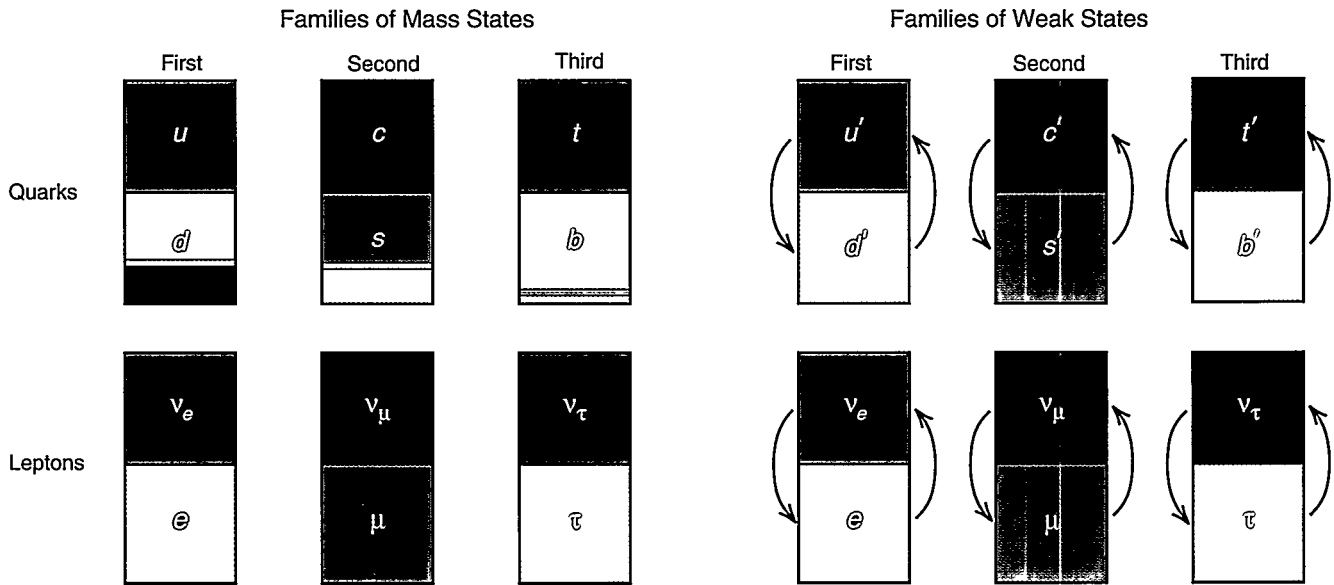
Figure 6. A Comparison of Mass States and Weak States in the Standard Model

The mass states (colored squares at left) and the weak states (colored squares at right) are two alternative descriptions of the spin-1/2 particles of the Standard Model. Here, the first, second, and third families of weak states are represented by colors: the greens, the purples, and the blues, respectively. By convention, all mixing among the quarks is placed in the lower half of the mass state quark doublets. Thus, the mass states d, s, and b are shown as mixtures of the particular colors that represent the quark weak states d' s', and b'. For example, the mass state d is mostly green but contains a purple stripe whose area represents the fraction of d in the weak state s', and so forth. Most of the quark mixing occurs between the first and second families. The mass and weak states for the quarks in the upper half of the doublets are equivalent: $u = u'$, $c = c'$, and $t = t'$. In the Standard Model, there is no mixing among the leptons, and so the lepton weak states and mass states are identical.

which must be very large to account for the short range and the reduced strength of the weak force relative to the electromagnetic force.

The same general mechanism is assumed to explain the masses of the quarks and leptons, but the theory has so many undetermined constants that experiment rather than theory is required to determine the masses. A theory of masses for spin-1/2 particles has yet to be found. Whatever the solution, it must give different masses to the clones in each family, because as can be seen in going from left to right along any row in Figure 1, the three families form a mass hierarchy from light to heavy. That is a tantalizing pattern with no explanation.

There are more mysteries surrounding the states defined by the weak symmetries (those shown in Figure 5). Why do quarks and leptons fall naturally into distinct families? Are these two types of particles related to each other in some way that is not yet apparent but that is anticipated in the Grand Unified Theories? Why are there three different families with exactly the same properties under the weak force? And why do they have different masses? Here, the Grand Unified Theories are no guide at all.

A related mystery is the one mentioned at the beginning of this section—the "nonalignment" between

the different quark states. Experiment shows that the quark states of definite mass (shown in Figure 1) are not the same as the quark states that make up the weak doublets. (Recall that the quark weak states have been labeled with primes.) The weak force seems to have a kind of skewed vision that produces and acts on quarks that are mixtures of the mass states from the different families. Equivalently, the symmetry-breaking mechanism that gives particles their masses mixes the quark clones in the weak families to create mass states.

Figure 6 stresses this point. Each family of weak states is denoted by a different color (green, purple, and blue), and the mass states are shown as mixtures of weak states (mixed colors). Areas of color represent the fraction of a mass state that is in a particular weak state. Notice that most of the quark mixing occurs between the first two families. The exact amounts of mixing cannot be derived from theory; instead, they are determined experimentally and included in the Standard Model as arbitrary parameters.[4] Notice also that, by convention, all the mixing is placed in the lower half of the quark doublets (the d, s, and b quarks are mixtures of d', s', and b').[5] Therefore, the weak and the mass states for the quarks in

Lepton Mass States

Lepton Weak States

| First | Second | Third | First | Second | Third |
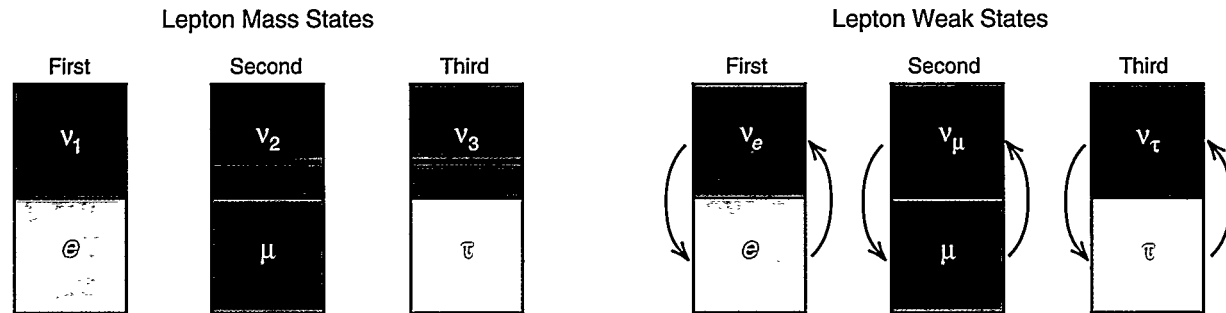
Figure 7. Lepton Mass States and Weak States for Nonzero Mixing among the Leptons
If neutrinos have mass and there is mixing among the leptons as there is among the quarks, all the mixing can be placed among the neutrinos, the neutral components of the weak doublets. Compare these lepton states with those in Figure 6. Although there is no mixing among the leptons in the Standard Model, present oscillation data suggest that such mixing may indeed occur. However, the pattern of mixing among the leptons is an open question. This figure suggests one possible pattern (shown by the color mixtures), which involves mainly the second and third families.

the upper half of the doublets are equivalent: $u = u'$, $c = c'$, and $t = t'$. It should be stressed, however, that no matter which way one views the mixing, the quark states that transmute into each other through the action of the W (red arrows) are *always* the members of the weak doublets.

The mixing that results from the nonalignment between mass and weak states is a natural outcome of the symmetry-breaking mechanism through which particles acquire mass in the Standard Model. According to the Higgs mechanism, the ground state, or lowest-energy state, has no physical particles (it is called "the vacuum"), but it contains an everpresent background of virtual, spin-zero Higgs particles. That background interacts with the quarks, leptons, and gauge bosons and provides a "drag" on them, which we observe as rest mass. The Higgs particles are weak doublets, and the background of virtual Higgs particles, by definition, has a nonzero value of weak charge. When the quarks and leptons in

the weak doublets "interact" with the Higgs background and acquire mass, the resulting states of definite mass do not conserve the weak charge and thus break the symmetry. Indeed, they cease being the states in the weak doublets. In the most general version of this symmetry-breaking, mass-generating scheme, the particles that acquire definite masses through interaction with the Higgs background are mixtures of the weak states from the different weak families. Indeed, the quarks in the Standard Model follow this most general scheme. The Higgs mechanism thus causes the mismatch between the quark weak states and mass states.

Since the leptons also acquire mass through the Higgs mechanism, one might expect to find a similar type of mixing among the lepton weak states and mass states. So far, experiments have not confirmed that expectation, and the Standard Model holds that the lepton mass states and weak states are essentially identical. The weak force always appears to act on the weak doublets within a family, and there is no mixing of weak states through the Higgs mechanism in the lepton sector. Consequently, one can define a quantity called lepton-family number that is conserved by all weak interactions involving the leptons. (Lepton-family numbers and the corresponding conservation laws are discussed later in this article.)

Why is there mixing among the quarks and not among the leptons? In the Standard Model, this difference follows directly from the assumption that all three neutrinos have the same mass, namely, zero. The mathematical argument is given in the sidebar "Family Mixing and the Origin of Mass" on page 72.

But as we said earlier, there is no fundamental principle that keeps the neutrinos massless. If they have small masses and acquire those masses through the Higgs mechanism, the mass states would likely be mixtures of the weak states. The lepton mass states would then change to look like those in Figure 7, in which the neutrino mass states ν_1, ν_2, and ν_3 are related to the three weak states ν_e, ν_μ, and ν_τ by a set of mixing parameters analogous to those relating the quark weak states to the quark mass states.

Mixing among the leptons would allow processes that violate lepton-family number, but because neutrinos have such small masses, we would expect most of those processes to be barely detectable. In fact, in a particular range of masses and mixings, the *only* example of lepton-family mixing that is accessible to measurement is neutrino oscillation, the spontaneous periodic change from one weak family to another as the neutrino propagates freely through space.

---

[4]The amounts of mixing determined from experiment become the numbers in the famous CKM matrix (named after Cabibbo, Kobayashi, and Maskawa), the unitary matrix that rotates the complete set of quark mass states into the complete set of quark weak states or vice versa.

[5]The freedom to put all the mixing in one-half of a weak isospin doublet depends on the fact that the weak force always acts between the two members of a weak doublet.

Look where we have arrived. We are saying that mixing among the leptons is a natural extension of the Standard Model if neutrinos have mass and that the most likely place to observe the mixing is in the peculiar manifestation of quantum mechanics known as neutrino oscillation. Furthermore, since oscillations can only occur if the neutrino types have different masses, direct observation of neutrino oscillations would reveal the relative sizes of the neutrino masses. No wonder that physicists have been searching for this phenomenon for well over two decades.

We will turn to the theory and detection of neutrino oscillations and examine how two types of information—neutrino masses and the amount of mixing across families—can be determined from oscillation data. But first, we will backtrack to the quarks and explain how mixing works.

Mixing among the Quarks

Consider ordinary neutron beta decay and suppose we had no idea of the difference between the weak states and the mass states. A neutron transforms into a proton, and an electron and an electron antineutrino are created in the decay process,

$$n \rightarrow p + e^- + \bar{\nu}_e \ . \qquad (1)$$

The neutron is made of the triplet of quark mass states udd, and the proton is made of the triplet of quark mass states uud. At the quark level, the change of a neutron to a proton looks like the transmutation of a down quark to an up quark $d \rightarrow u$ (refer to Figure 3a). However, when the strength (effective coupling) of the force responsible for neutron beta decay is measured, it is found to be 4 percent smaller than the strength of the force responsible for muon beta decay (refer to Figure 3b). But these are just two different examples of the charged-current weak force, and theory says that the strength of the force should be

Mass states

First   Second

Weak states

First   Second

$$\begin{pmatrix} d \\ s \end{pmatrix} = \begin{pmatrix} \cos\theta_c & -\sin\theta_c \\ \sin\theta_c & \cos\theta_c \end{pmatrix} \begin{pmatrix} d' \\ s' \end{pmatrix}$$

mass   mixing   weak
states   matrix   states

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}$$

weak   mixing   mass
states   matrix   states
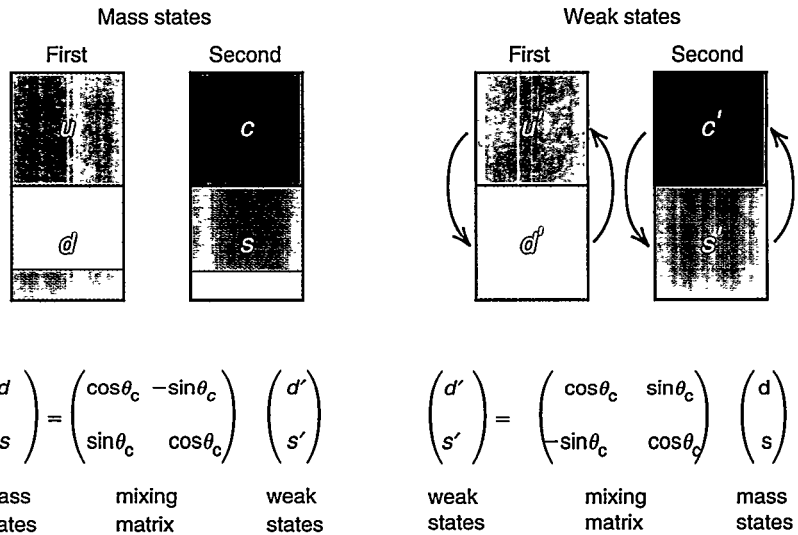
Figure 8. Two-Family Mixing among the Quarks

The quark weak states and mass states are like two alternative sets of unit vectors in a plane (see diagram at right) that are related to each other by the rotation through an angle θ_c. In this analogy, one mixing matrix is just a rotation matrix that takes, say, the mass coordinates d and s into the weak-force coordinates d' and s'; its inverse is the rotation through the angle $-\theta_c$ that takes the weak coordinates into the mass coordinates.
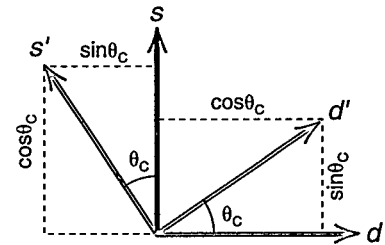
identical in the two processes.

Where did the missing strength of the weak force go? It turns out to be "hiding" in the beta decay of the lambda particle (Λ):

$$\Lambda \rightarrow p + e^- + \bar{\nu}_e \ . \qquad (2)$$

The lambda (uds) differs from the neutron (udd) by having a strange quark replace a down quark. The lambda decays to a proton because the strange quark transforms into an up quark, $s \rightarrow u$. Lambda beta decay is thus analogous to neutron beta decay, and the sum of the strengths for lambda and neutron beta decays equals the strength for muon beta decay.

Why is this so? The answer is mixing—the fact that the quark mass states that appear in the neutron and the lambda are mixtures of the quark weak

states. The mathematics of this mixing is interesting not only for tracking down the missing 4 percent, but also because it has the same form as the mixing that causes neutrino oscillations. For simplicity, we will consider mixing between the first two families of quarks only, which accounts for most of the mixing among the quarks anyway.

Figure 8 shows the quark weak states and the quark mass states in the two-family picture. Underneath the families of mass states, the 2×2 rotation matrix is shown, which rotates the weak states d' and s' into the mass states d and s; the inverse transformation is shown under the families of weak states. In this quantum mechanical world, the quark mass states s and d are one complete description of the quarks with electric charge $Q = -1/3$. The quark weak states s'

Neutron Decay

$$n \longrightarrow p + e^- + \bar{v}_e$$
$$(udd)$$
$$\hookrightarrow d = \cos\theta_c d' - \sin\theta_c s'$$
$$\downarrow$$
$$(uud)$$
$$p + e^- + \bar{v}_e$$

Lambda Decay

$$\Lambda \longrightarrow p + e^- + \bar{v}_e$$
$$(usd)$$
$$\hookrightarrow s = \sin\theta_c d' + \cos\theta_c s'$$
$$\downarrow$$
$$(uud)$$
$$p + e^- + \bar{v}_e$$

Initial state   Final state

Neutron (u d d) Proton

$\cos\theta_c$ W^- e^- Electron

$\bar{v}_e$ Electron antineutrino

$P(n \to p + e^- + \bar{v}_e) \propto \cos^2\theta_c$

The neutron decay amplitude is proportional to $\cos\theta_c$, the amplitude of the state d' in the d quark. The decay probability is proportional to the square of the amplitude:

Initial state   Final state

Lambda (u d s) Proton

$\sin\theta_c$ W^- e^- Electron

$\bar{v}_e$ Electron antineutrino

$P(\Lambda \to p + e^- + \bar{v}_e) \propto \sin^2\theta_c$

The lambda decay amplitude is proportional to $\sin\theta_c$, the amplitude of the state d' in the s quark. The lambda decay probability is proportional to the square of the amplitude.

Figure 9. Neutron and Lambda Beta Decay in the Two-Family Picture

In beta decay, the neutron transforms into a proton through the transition $d \to u$, and the lambda transforms into a proton through the transition $s \to u$. However, in both cases, the W acts between members of the quark weak isospin doublets in the first family, that is, the W causes the transition $d' \to u$. So, only the fraction of the d in the state d' takes part in neutron decay, and only the fraction of the s in the state d' takes part in lambda decay. The multicolored lines for d and s show their fractional content of d' (green) and s' (purple).

and d' are an alternative description, and the two sets of states are like two independent sets of orthogonal unit vectors in a plane that are related to each other by a rotation through the angle θ_c, also called the mixing angle. Thus, the weak states can be described as linear combinations of the mass states, and conversely, the mass states can be described as linear combinations of the weak states.

The phenomenon of mixing, while perhaps nonintuitive, emerges naturally from the fundamental tenet of quantum mechanics that particles have wavelike properties. Like sound and light waves, matter waves, or quantum mechanical states, can add together to form a coherent linear superposition of waves. We will see later that the neutrinos produced in weak processes may like-

wise be linear superpositions of different neutrino mass states, and those mass states, or matter waves, can generate the interference patterns that we call oscillations.

But first, let us track down the 4 percent decrease in the expected rate of neutron beta decay. Figure 8 also shows the weak quark doublets that transform into each other through interaction with the W in the two-family picture. The weak doublets are not (u, d) and (c, s), but rather (u, d') and (c, s'). Now, consider the beta decay of the lambda and the neutron. As shown in Figure 9, both decays involve the transition $d' \to u$. The d quark in the neutron and the s quark in the lambda are mass states that contain a fraction of d'. The compositions of these mass states are given by

$$|d\rangle = \cos\theta_c|d'\rangle - \sin\theta_c|s'\rangle \ ;$$
$$|s\rangle = \sin\theta_c|d'\rangle + \cos\theta_c|s'\rangle \ . \tag{3}$$

Figure 9 also illustrates that the transition amplitude for a neutron to turn into a proton (that is, for a d quark to turn into a u quark) is proportional to $\cos\theta_c$, or the amplitude of the d quark that is in the state d'. Similarly, the transition amplitude for the lambda to turn into a proton (that is, for an s quark to change into a u quark) is proportional to $\sin\theta_c$, the amplitude of the s quark that is in the state d'.

The rate of neutron beta decay is proportional to the square of that transition amplitude and is thus proportional to $\cos^2\theta_c$. The rate of lambda beta decay is proportional to $\sin^2\theta_c$. The sum of the rates for the two processes equals the rate for the transi-

CKM   CKM   CKM   CKM

d W $\bar{d}$ $\bar{d}$ W d

K^0 u,c,t u,c,t $\bar{K}^0$ u,c,t u,c,t K^0

$\bar{s}$ W s s W $\bar{s}$

Particle   CKM   CKM   Antiparticle   CKM   CKM   Particle
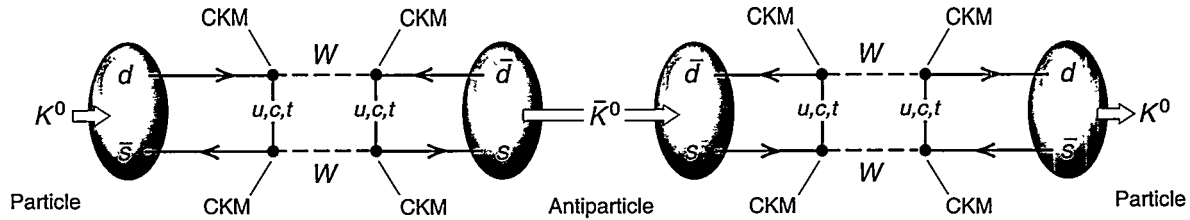
Figure 10. Oscillation of the Neutral Kaons

The neutral kaon K^0 ($s\bar{d}$) can transform into its antiparticle $\bar{K}^0$ ($\bar{s}d$) and back again, in each case through the four weak-interaction vertices shown above. The CKM matrix at each vertex indicates that the transitions mediated by the W are between members of the weak doublets, and they can proceed only because the quark mass states in the neutral kaons contain mixtures of the weak states d', s', and b'.

tion of d' into u. That rate is the same for transitions between all weak doublets, including the leptonic transition $\mu \rightarrow \nu_\mu$ in muon beta decay shown in Figure 3(b).

The mixing angle θ_c for these first two families is called the Cabibbo angle, and it has been determined from experiment. The 4 percent decrease in the rate of neutron beta decay relative to muon beta decay provides a measure of that angle: $1 - \cos^2\theta_c = \sin^2\theta_c \approx 0.04$. And that decrease is made up for by the rate of lambda beta decay. The measured value for $\sin^2\theta_c$ is 0.22.

In the Standard Model, the mixing between the quark weak and mass states occurs among the three families, not just two, and the amounts of mixing appear in the famous CKM matrix, the 3×3 unitary mixing matrix for the three quark families that is analogous to the 2×2 rotation matrix in Figure 8. In the Standard Model, the mismatch between quark mass states and weak states is responsible for all processes in which quarks transmute across family lines. Among those processes is the oscillation between the neutral kaon, K^0 ($s\bar{d}$), and its antiparticle, $\bar{K}^0$ ($\bar{s}d$). The kaons periodically change from particle to antiparticle during free flight in space. Figure 10 shows how oscillations can come about as the quark mass states composing the kaons interact through the W. The quarks transmute across family lines because they are mass states, each a mixture of weak states from all three weak families. Just as the mixing

between the quark weak and mass states results in the oscillation of the neutral kaon into its antiparticle, the oscillation of one neutrino flavor into another is possible only if there is mixing between lepton weak and mass states.

Nonmixing among Leptons and Lepton-Number Conservation Laws

To recap what we discussed earlier, in the usual version of the Standard Model, there is no mixing among the leptons. Because the three neutrinos are assumed to have the same mass (namely, zero), the lepton version of the CKM mixing matrix for quarks is the identity matrix. Thus, the mass states and weak states are equivalent, and there is no mechanism to produce reactions that will cross family lines. As with the quarks, the weak force always acts between the members of a weak doublet and simply transforms a muon into a muon neutrino and vice versa, or allows similar transformations for the other lepton families. A further assumption in the Standard Model is that, although electrically neutral, the left-handed neutrino and the right-handed antineutrino are distinct particles and cannot transmute into each other.

These theoretical assumptions lead directly to two types of lepton-number conservation laws: one for total lepton number (the number of leptons minus

the number of antileptons) and the other for individual-lepton-family number (the number of leptons minus the number of antileptons in a particular lepton family). Although these laws can be viewed as predictions of the Standard Model, they were deduced empirically a decade before the Standard Model was formulated. Let us review the relevant leptonic reactions and methods of interpretation because the same reactions are now being used to detect neutrino oscillations and to search for the consequences of nonzero neutrino masses.

Conservation of Total Lepton Number. The primary sources of neutrinos in cosmic-ray- and accelerator-based neutrino experiments are pion and muon decays. Pions[6] come in three charge states, the π^+, π^-, and π^0.

Shortly after they are produced through the strong force, the charged pions decay into muons through the weak force:

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \; ;$$

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \; . \tag{4}$$

[6]The pion is a massive spin-0 particle made of quark-antiquark pairs from the first family. It is a carrier, or mediator, of the residual strong force that binds neutrons and protons inside nuclei. Pions are produced, or "boiled off," in great numbers when nuclei are bombarded by energetic protons. Yukawa predicted the existence of this particle in the 1930s. When the muon, which is slightly less massive than the pion, was discovered in cosmic rays in 1937, it was at first mistakenly identified as Yukawa's particle.
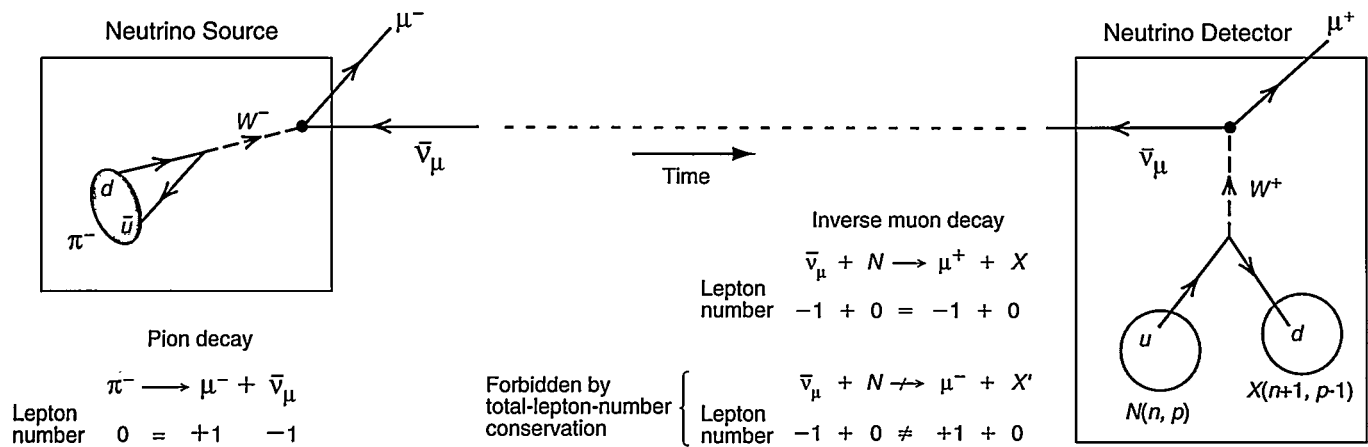
Figure 11. Test of Lepton-Number Conservation

At left is a neutrino source consisting of muon antineutrinos $(L = -1)$ from pion decay. If total lepton number is conserved, then as shown in the figure, those antineutrinos should interact with matter through inverse muon decay and produce antimuons $(L = -1)$. They should never produce muons because that reaction would change the total lepton number by two units. Shown in the figure are the lepton numbers for pion decay and inverse muon decay as well as the reaction forbidden by total-lepton-number conservation.

The positive pion will decay to the antimuon and the negative pion to the muon because electric charge must be conserved. The law of total-lepton-number conservation says that the number of leptons minus the number of antileptons must not change in any reaction. To formalize this law, every particle is assigned a lepton number L. By convention, the negatively charged leptons are called leptons and assigned a lepton number of $+1$, and their positively charged counterparts are called antileptons and are assigned a lepton number of -1 (see Table I). Because quarks are not leptons, they are assigned a lepton number of zero.

Since the pion is also not a lepton (lepton number $L = 0$), its decay must produce one lepton and one antilepton $(L = 1 - 1 = 0)$. Thus a ν_μ (lepton) is created with the μ^+ (antilepton), or a $\bar{\nu}_\mu$ (antilepton) is created with the μ^- (lepton). Conservation of total lepton number can easily be checked in all the processes shown in Figures 3 and 4.

How can one prove that the neutrino and antineutrino have different lepton numbers? How can one show that, for example, the neutrino from π^+ decay has lepton number $+1$, like the muon, whereas the antineutrino from π^- decay has lepton number -1, like the antimuon? The test requires detecting the interaction of those neutrinos with matter. As shown in Figure 11, the antineutrino from π^- decay has lepton number -1 if it produces an antimuon $(L = -1)$—and never a muon—when interacting with matter. Likewise, the neutrino from π^+ decay has lepton number $+1$ if it produces a muon—never an antimuon.

Indeed, these tests have been performed, and conservation of total lepton number holds to a very high level of precision. (However, if neutrinos have a small nonzero mass and furthermore if they acquire that mass through what is called a Majorana mass term, neutrinos would be their own antiparticles. They

Table I. Lepton Numbers and Lepton-Family Numbers

Particle	Lepton Number L	Electron-Family Number L_e	Muon-Family Number L_μ	Tau-Family Number L_τ
e^-	$+1$	$+1$	0	0
ν_e	$+1$	$+1$	0	0
e^+	-1	-1	0	0
$\bar{\nu}_e$	-1	-1	0	0
μ^-	$+1$	0	$+1$	0
ν_μ	$+1$	0	$+1$	0
μ^+	-1	0	-1	0
$\bar{\nu}_\mu$	-1	0	-1	0
τ^-	$+1$	0	0	$+1$
ν_τ	$+1$	0	0	$+1$
τ^+	-1	0	0	-1
$\bar{\nu}_\tau$	-1	0	0	-1

Figure 12. Is the Muon Neutrino the Same Particle as the Electron Neutrino?
Ordinary muon decay is shown in (a). At one weak-interaction vertex, a muon transmutes into a muon neutrino and emits a W^-, and at the second, the W^- decays into an electron and an electron antineutrino. Two neutrinos are produced, one associated with the muon and the other with the electron. (b) If the muon neutrino were the same as the electron neutrino, then the muon could decay to an electron through two weak-interaction vertices. At one vertex, the muon transforms into a neutrino and emits a W^-; at the second vertex, that same neutrino absorbs a W^- and transforms into an electron. To conserve energy and momentum, the (virtual) W^- radiates a photon. Thus, muon decay produces an electron and a gamma ray, but no neutrinos are emitted. In other words, the process $\mu^- \rightarrow e^- + \gamma$ could occur if the muon neutrino were the same as the electron neutrino.

(a) Initial state — Final state
Muon μ^- — ν Neutrino
W^-
e^- Electron
$\bar{\nu}_e$ Electron antineutrino

(b) Initial state — Final state
Muon μ^- — Neutrino ν — Electron e^-
W^- W^-
Gamma γ
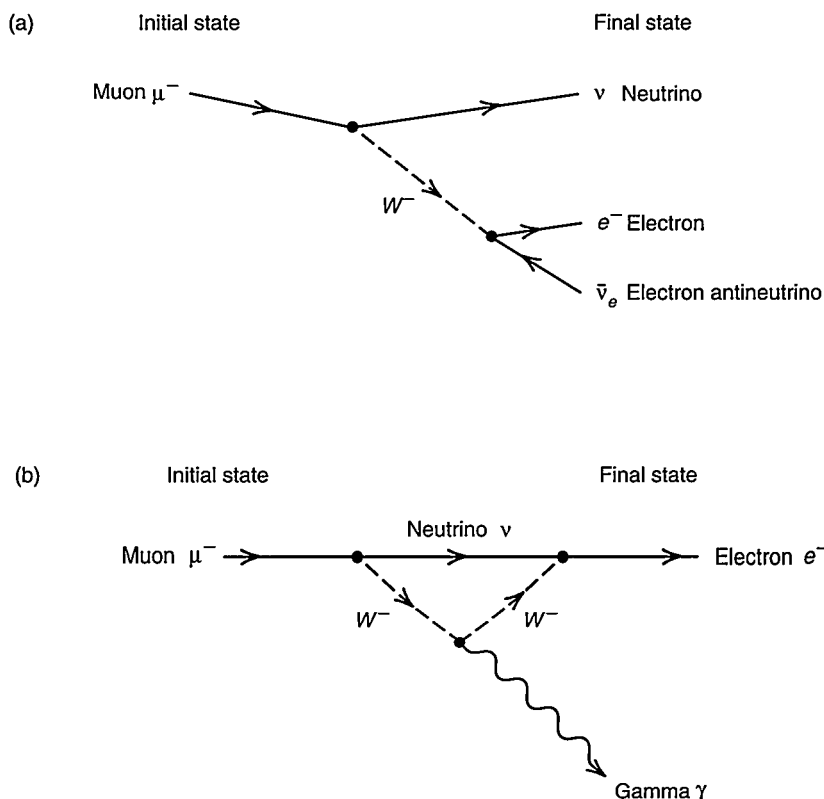
might induce, at some low rate, reactions that would change total lepton number. This possibility will be discussed later in the text.)

Conservation of Lepton-Family Number. One might also wonder how it was shown that the muon neutrino is really distinct from the electron neutrino and that distinct lepton families are under the weak force. Those discoveries came from studies of muon decay. In the late 1940s, the muon was observed to decay into an electron emitted with a spectrum of energies. As in ordinary beta decay, a spectrum of electron energies rather than a single energy means that the decay must yield three particles in the final state.

However, only the electron revealed its presence, so the two unidentified particles ($\mu \rightarrow e + ? + ?$) had to be electrically neutral. It was also observed that the rates of muon decay and muon capture by nuclei were very similar to the rates for beta decay and electron capture. The same weak force

outlined in Fermi's theory of beta decay seemed to be at work, and so the mechanism of muon decay was believed to be entirely analogous to that of neutron beta decay.

At that time, the local symmetry of the weak force was not known, but Fermi's theory did place particles in pairs that transformed into each other under the weak force. It was therefore assumed that the weak force transformed the muon into a neutral particle of some kind, perhaps the neutrino, and that, to conserve charge, an electron and an antineutrino were produced as in ordinary neutron beta decay:

$$\mu \rightarrow \nu + e + \bar{\nu}_e \ . \qquad (5)$$

Then, in the 1950s, theorists considered the possibility that a massive gauge boson (like the W) mediated the weak force, in which case the muon could decay to an electron through the two processes shown in Figure 12. The latter process involves not only the exchange of a virtual W but also the

exchange of a virtual neutrino that couples to both the electron and the muon. In other words, the muon transmutes into a neutrino, and then that same neutrino transmutes into an electron. Because there are three interaction vertices in the diagram for $\mu^- \rightarrow e^- + \gamma$, two weak and one electromagnetic, the rate for this second mode would be small but still observable, about 10^{-5} of the total decay rate of the muon. This decay mode, however, has never been observed. The MEGA (muon to electron plus gamma) experiment, currently nearing completion at Los Alamos, has put the most stringent upper limit on the rate of this process so far. It is less than 4×10^{-11} of the total muon-decay rate.

The absence of $\mu^- \rightarrow e^- + \gamma$ is a clue that there are two neutrino flavors—one strictly associated with the electron; the other, with the muon. In muon beta decay, for example, a muon transforms into a *muon* neutrino, and an electron and its antineutrino are created to conserve charge:

$$\mu^- \to \nu_\mu + e^- + \bar{\nu}_e \ . \qquad (6)$$

Likewise, in π^+ decay, the neutrino created with the antimuon is a *muon* neutrino ($\pi^+ \to \mu^+ + \nu_\mu$), not an electron neutrino. Thus, a second lepton family was thought to exist.

The conjecture of two neutrino flavors was tested by Leon Lederman, Mel Schwartz, and Jack Steinberger, who designed an ingenious experiment—analogous to the one illustrated in Figure 11—at the Brookhaven 30-giga-electron-volt (GeV) proton accelerator. As in most accelerator-neutrino experiments, a pulsed beam of protons is directed at a target, where they produce a myriad of pions that rapidly decay into muons and neutrinos.

In this case, the experimenters found a way to tailor a narrow beam of high-energy neutrinos from a much wider distribution. They allowed these high-energy neutrinos to pass through a huge spark chamber containing 10 tons of aluminum plates in parallel stacks separated by narrow gaps. A neutrino entering the spark chamber could interact with an aluminum nucleus and produce a high-energy muon or electron. Either one would leave an ionization track in the gas between the plates, and if the plates were charged, they would discharge along that track and create a trail of bright sparks that could easily be photographed. The experiment produced a total of 29 spark-chamber photographs containing long, straight tracks that started from within the spark chamber and were characteristic of an energetic muon. The erratic, staggered tracks that would be produced by the much lighter electron were essentially absent. Thus, the neutrino produced in π^+ decay could transform into a muon but not into an electron.

These results supported the idea of two independent neutrino flavors and led the way for establishing separate conservation laws for two new quantum numbers, muon-family number and electron-family number (refer again to

Table II. Decays Forbidden by Lepton-Family-Number Conservation Laws

$$\mu^+ \to e^+ + \gamma$$

$$\mu^+ \to e^+ + e^- + e^+$$

$$\mu^- + N(n, p) \to e^- + N(n, p)$$

$$\mu^- + N(n, p) \to e^+ + N(n + 2, p - 2)$$

$$\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu$$

Table I). These laws are analogous to the conservation laws of total lepton number except that they apply separately to the electron, electron neutrino, and their antiparticles on the one hand and to the muon, muon neutrino, and their antiparticles on the other.

To conserve muon-family number, a muon can turn into a muon neutrino—never into a particle with a muon number of zero. Similarly, to conserve electron number, an electron can turn into an electron neutrino; it cannot turn into a particle with an electron number of zero. Once the tau, the charged lepton of the third family, was discovered, the tau neutrino was assumed to exist, and tau-family number and its conservation were postulated.

At the beginning of this section, we stated that strict separation between the lepton families is implied by the gauge symmetry of the weak force, combined with the assumption that the three neutrinos are massless. But this assumption always seemed to rest on shaky ground. More important, new forces could exist, even weaker than the weak force, that have yet to be seen but that allow leptons to transmute across family lines. Consequently, there have been many searches for various "forbidden" reactions such as those listed in Table II. Searches for violations of the Standard Model have mostly reported null results. The exception is the LSND experiment, which reports that muon antineutrinos can oscillate into electron antineutrinos with a probability of about 0.3 percent

(averaged over the experimental energy and distances).

Neutrino Oscillations

The first suggestion that free neutrinos traveling through space might oscillate, that is, periodically change from one neutrino type to another, was made in 1957 by Bruno Pontecorvo. Gell-Mann and Pais had just shown how quantum mechanical interference would allow the neutral kaon K^0 ($s\bar{d}$) and its antiparticle $\bar{K}^0$ ($\bar{s}d$) to oscillate back and forth because the quark mass states are mixtures of weak states. Pontecorvo noted very briefly that, if the neutrino had mass and if total lepton number were not conserved, the neutrino could imitate the neutral kaon, oscillating between particle and antiparticle as it travels through empty space. This possibility would have implied that the neutrino is a massive Majorana particle with no definite distinction between particle and antiparticle forms.

Although very interesting and still relevant today, Pontecorvo's suggestion was not explored in 1957 because Lee and Yang's theory of the massless two-component neutrino was just gaining acceptance. This theory helped explain why parity was maximally violated in nuclear beta decay. The existence of a left-handed neutrino, distinct from the right-handed antineutrino by having the opposite lepton number, was a crucial postulate (see the box "Parity Nonconservation and the Massless Two-

Component Neutrino" on page 32). In that theory, particle-antiparticle oscillations could not occur.

Solar Neutrinos. In 1963, after Lederman, Steinberger, and Schwartz showed that there were two distinct flavors of neutrino, the idea of oscillation between electron neutrinos and muon neutrinos surfaced for the first time. This possibility requires mixing across the lepton families as well as nonzero neutrino masses. In 1969, it was decided that the idea of neutrino oscillation was worth testing. The Sun is known to drench us with low-energy electron neutrinos that are produced in the thermonuclear furnace at its core, as shown in Figure 13(a). By using standard astrophysics models about stellar processes and the observed value of the Sun's luminosity, theorists can predict the size of the neutrino flux. But measurements of the solar-neutrino flux present an intriguing puzzle: A significant fraction of those electron neutrinos apparently disappear before reaching our terrestrial detectors. Ray Davis made the first observation of a neutrino shortfall at the Homestake Mine in South Dakota, and all experiments since have confirmed it. Today, the most plausible explanation of the solar-neutrino puzzle lies in the oscillation of electron neutrinos into other types of neutrinos. Although the measured shortfall is large and the expected amplitude for neutrino oscillations in a vacuum is small, neutrino oscillations can still explain the shortfall through the MSW effect.

Named after Mikheyev, Smirnov, and Wolfenstein, the MSW effect describes how electron neutrinos, through their interactions with electrons in solar matter, can dramatically increase their intrinsic oscillation probability as they travel from the solar core to the surface. This matter enhancement of neutrino oscillations varies with neutrino energy and matter density. The next generation of solar-neutrino experiments is specifically designed to explore whether the electron neutrino deficit has the energy dependence predicted by the MSW effect (see the articles "Exorcising Ghosts" on page 136 and "MSW" on page 156).

Atmospheric Neutrinos. In 1992, another neutrino deficit was seen—this time in the ratio of muon neutrinos to electron neutrinos produced at the top of the earth's atmosphere. When high-energy cosmic rays, mostly protons, strike nuclei in the upper atmosphere, they produce pions and muons, which then decay through the weak force and produce muon and electron neutrinos. The atmospheric neutrinos have very high energies, ranging from hundreds of million electron volts (MeV) to tens of giga-electron-volts, depending on the energy of the incident cosmic ray and on how this energy is shared among the fragments of the initial reaction. As shown in Figure 13(b), the decay of pions to muons followed by the decay of muons to electrons produces two muon neutrinos for every electron neutrino. But the measured ratio of these two types is much smaller (see the article "The Evidence for Oscillations" on page 116). The oscillation of muon neutrinos into tau neutrinos appears to be the simplest explanation.

Accelerator Neutrinos. The lone accelerator-based experiment with evidence for neutrino oscillations is LSND. This experiment uses the high-intensity proton beam from the linear accelerator at the Los Alamos Neutron Science Center (LANSCE) to generate an intense source of neutrinos with average energies of about 50 MeV. In 1995, the LSND collaboration reported positive signs of neutrino oscillations. An excess of 22 electron antineutrino events over background was observed. They were interpreted as evidence for the oscillation of muon antineutrinos, into electron antineutrinos (see Figure 13c). The muon antineutrinos had been produced at the accelerator target through antimuon decay-at-rest. As in the experiments described earlier to study electron-family-number and muon-family-number conservation laws, the electron antineutrino was detected through its charged-current interaction with matter, that is, through inverse beta decay.

Recently, members of the LSND collaboration reported a second positive result. This time, they searched for the oscillation of muon neutrinos rather than muon antineutrinos. The muon neutrinos are only produced during pion decay-in-flight, before the pions reach the beam stop. Therefore, these neutrinos have a higher average energy than the muon antineutrinos measured in the earlier experiment. The muon neutrinos were observed to turn into electron neutrinos at a rate consistent with the rate for antineutrino oscillation reported earlier. Since the two experiments involved different neutrino energies and different reactions to detect the oscillations, the two results are indeed independent. The fact that the two results confirm one another is therefore most significant. The complete story of LSND can be found in the article "A Thousand Eyes" on page 92.

Each type of experiment shown in Figure 13, when interpreted as an oscillation experiment, yields information about the oscillation amplitude and wavelength. One can therefore deduce information about the sizes of neutrino masses and lepton-family mixing parameters. The specific relationships are explained in the next section.

The Mechanics of Oscillation

Oscillation, or the spontaneous periodic change from one neutrino mass state to another, is a spectacular example of quantum mechanics. A neutrino produced through the weak force in, say, muon decay, is described as the sum of two matter waves. As the neutrino travels through space (and depending on which masses are measured), these matter waves interfere with each other constructively or destructively. For example, the interference causes first the disappearance and

Figure 13. Three Types of Evidence for Neutrino Oscillations

(a) Solar neutrinos—a disappearance experiment. The flux of electron neutrinos produced in the Sun's core was measured in large underground detectors and found to be lower than expected. The "disappearance" could be explained by the oscillation of the electron neutrino into another flavor.

Sun

Earth

Primary neutrino source

$\sim 10^8$ kilometers

Underground V_e detector

$p + p \longrightarrow D + e^+ + V_e$
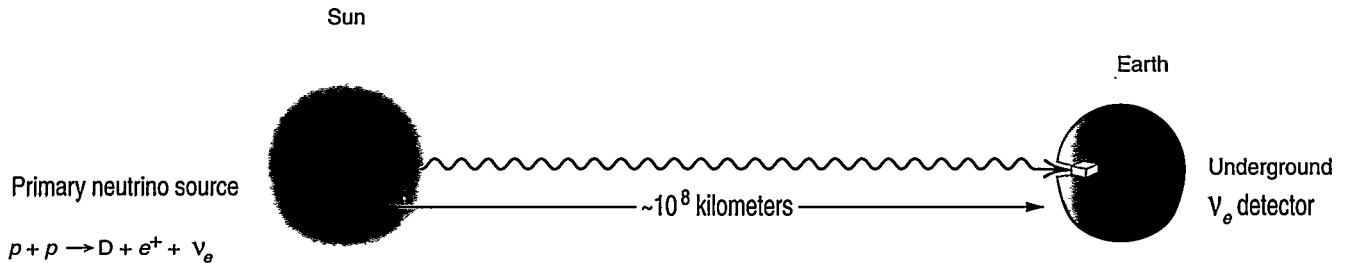
Other sources of neutrinos:

$e^- + {}^7Be \longrightarrow {}^7Li + V_e$

${}^8B \longrightarrow 2\,{}^4He + e^+ + V_e$
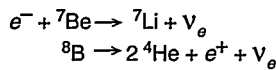
(b) Atmospheric neutrinos—a disappearance experiment. Collisions between high-energy protons and nuclei in the upper atmosphere can create high-energy pions. The decay of those pions followed by the decay of the resulting muons produces twice as many muon-type neutrinos (blue) as electron-type neutrinos (red). But underground neutrino detectors designed to measure both types see a much smaller ratio than 2 to 1. The oscillation of muon neutrinos into tau neutrinos could explain that deficit.

Cosmic-ray shower

π^0

π^+

μ^+

~ 30 kilometers

e^+

V_e

$\bar{V}_\mu$

V_μ

Underground V_e, $\bar{V}_e$, V_μ, $\bar{V}_\mu$ detector

Atmospheric neutrino source

$\pi^+ \longrightarrow \mu^+ + V_\mu$
$\quad\longrightarrow e^+ + V_e + \bar{V}_\mu$

$\pi^- \longrightarrow \mu^- + \bar{V}_\mu$
$\quad\longrightarrow e^- + \bar{V}_e + V_\mu$
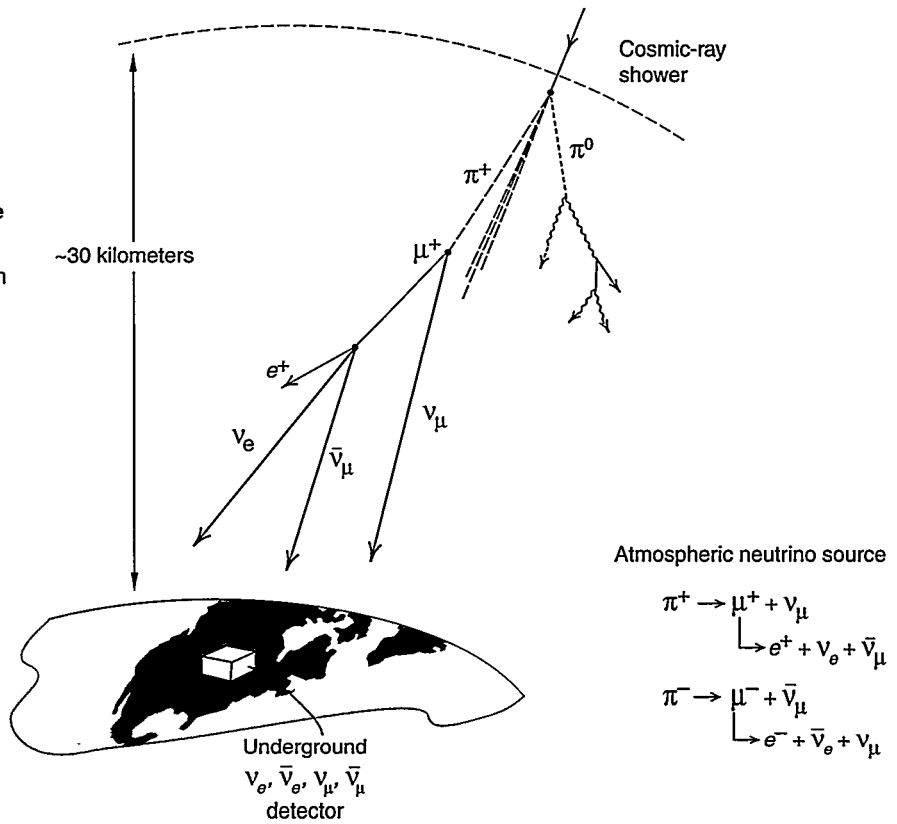
(c) LSND—an appearance experiment. Positive pions decay at rest into positive muons, which then decay into muon antineutrinos, positrons, and electron neutrinos. Negative pions decay and produce electron antineutrinos, but that rate is almost negligible. A giant liquid-scintillator neutrino detector located 30 meters downstream looks for the appearance of electron antineutrinos as the signal that the muon antineutrinos have oscillated into that flavor.

Proton beam

Pions

Water target

Muons and electrons

Copper beam stop

Neutrinos V_e, V_μ, and $\bar{V}_\mu$

$\bar{V}_e$ detector

30 meters
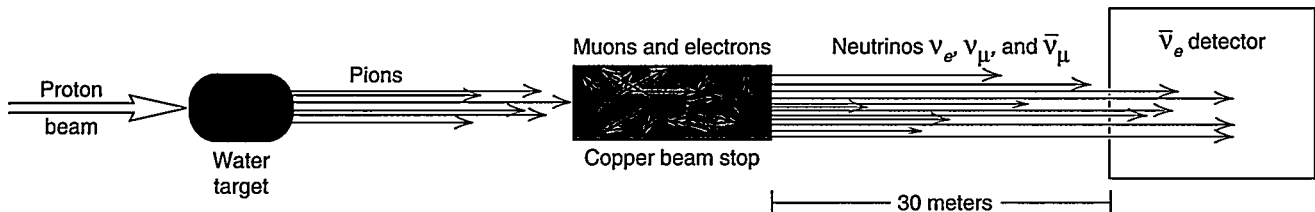
Figure 14. Neutrino Oscillations in the Two-Family Context

Mass states

First   Second

ν_1

Weak states

First   Second

ν_e

(a) Neutrino mass states and weak states. The weak states ν_e and ν_μ are shown as color mixtures of the mass states ν_1 (yellow) and ν_2 (red), and the mixing matrix that rotates ν_1 and ν_2 into ν_e and ν_μ is shown below the weak states. Each set of states is also represented as a set of unit vectors in a plane. The two sets are rotated by an angle θ relative to each other.

$$\begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}$$
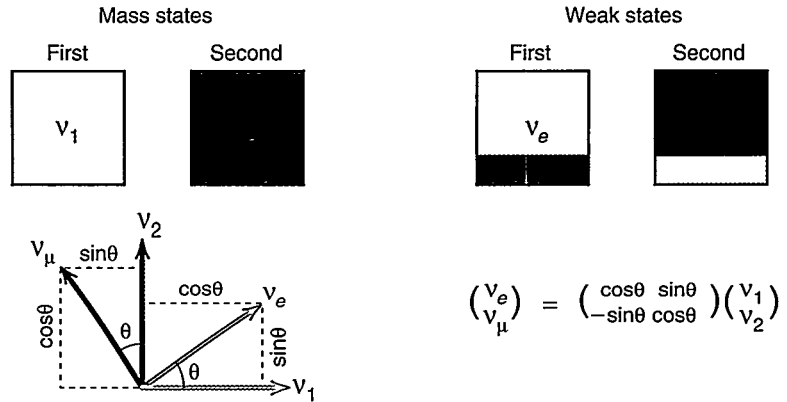
(b) Time evolution of the muon neutrino. The ν_μ is produced at $t = 0$ as a specific linear combination of mass states: $\nu_\mu = -\sin\theta\,\nu_1 + \cos\theta\,\nu_2$. The amplitude of each mass state is shown oscillating in time with a frequency determined by the energy of that mass state. The energies of the two states are different because their masses are different, $m_1 \neq m_2$. Each time the two mass states return to the original phase relationship at $t = 0$, they compose a pure ν_μ. At other times, the two mass states have a different phase relationship and can be thought of as a mixture of ν_μ and ν_e.

Pure ν_μ Pure ν_μ Pure ν_μ

$\cos\theta$ $-\sin\theta$ ν_2 ν_1

0   Time, t
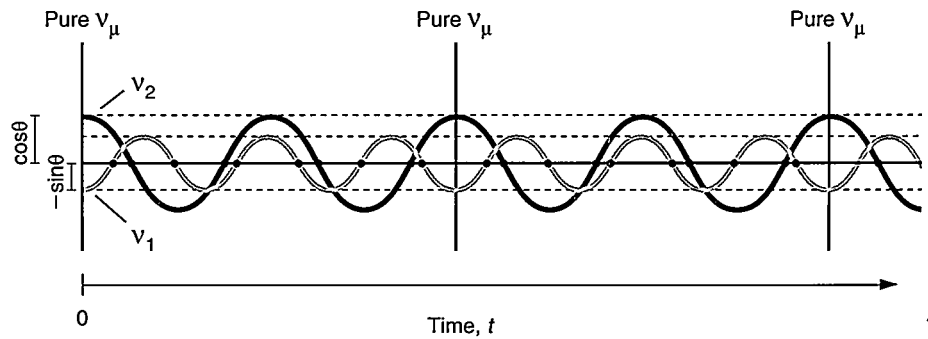
(c) Neutrino oscillation. Because the two mass components interfere with each other, the probability of finding a muon neutrino (purple) oscillates with distance from the source. The probability of finding an electron neutrino in its place also oscillates, and in the two-family approximation, the sum of the probabilities is always 1. The wavelength of this oscillation λ_{osc} increases as the masses of the two neutrinos get closer in value.

Source

ν_μ ν_e ν_e ν_e

ν_μ ν_μ ν_μ ν_μ ν_μ ν_μ

$P(\nu_\mu \rightarrow \nu_e) + P(\nu_\mu \rightarrow \nu_\mu)$

$\sin^2 2\theta$

$$\lambda_{osc} = \frac{2.5\,E_\nu}{\Delta m^2}$$

Distance, $x = ct$

☐ Probability that ν_μ has become ν_e ☐ Probability that ν_μ is still ν_μ
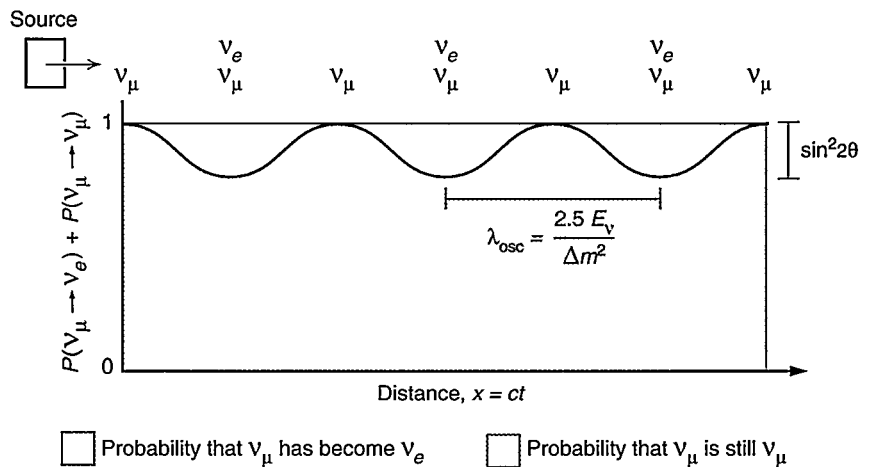
then the reappearance of the original type of neutrino. The interference can occur only if the two matter waves have different masses. Thus, the mechanics of oscillation start from the assumption that the lepton weak and mass states are not the same and that one set is composed of mixtures of the other set in a manner entirely analogous to the descriptions of the quark weak and mass states in Figure 8. In other words, there must be mixing among the leptons as there is among the quarks.

In the examples of quark mixing described earlier, the quarks within the composite particles (proton, neutron, lambda) start and end as pure mass states, and the fact that they are mixtures of weak states shows up through the action of the weak force. When a neutron decays through the weak force and the d quark transforms into a u, only a measurement of the decay rate reflects the degree to which a d quark is composed of the weak state d'. In contrast, in neutrino oscillation experi-

ments, the neutrinos always start and end as pure weak states. They are typically created through weak-force processes of pion decay and muon decay, and they are typically detected through inverse beta decay and inverse muon decay, weak processes in which the neutrinos are transmuted back to their charged lepton partners. Between the point of creation and the point of detection, they propagate freely, and if they oscillate into a weak state from a different family, it is not through the

action of the weak force, but rather through the pattern of interference that develops as the different mass states composing the original neutrino state evolve in time.

To see how the oscillation depends on the masses of the different neutrino mass states as well as the mixing angles between the lepton families, we limit the discussion to the first two families and assign the mixing to the electron neutrino and the muon neutrino (the halves of the lepton weak doublets with $I_3^W = 1/2$, as shown in Figure 5). Instead of expressing the mass states in terms of the weak states, as was done in Equation (3), we can use the alternate point of view and express the neutrino weak states $|\nu_e\rangle$ and $|\nu_\mu\rangle$ as linear combinations of the neutrino mass states $|\nu_1\rangle$ and $|\nu_2\rangle$ with masses m_1 and m_2, respectively (where we have assumed that m_1 and m_2 are not equal). Figure 14(a) illustrates this point of view. It shows how the weak states and mass states are like alternate sets of unit vectors in a plane that are related to each other by a rotation through an angle θ. The rotation, or mixing, yields the following relationships:

$$|\nu_e\rangle = \cos\theta|\nu_1\rangle + \sin\theta|\nu_2\rangle \;;$$
$$\tag{7}$$
$$|\nu_\mu\rangle = -\sin\theta|\nu_1\rangle + \cos\theta|\nu_2\rangle \;.$$

The mixing angle θ is the lepton analog of the Cabibbo mixing angle for the quarks. If θ is small, then $\cos\theta$ is close to 1, and the electron neutrino is mostly made of the state with mass m_1, whereas the muon neutrino is mostly made of the state with mass m_2. If the mixing angle is maximal (that is, $\theta = \pi/4$, so that $\cos\theta = \sin\theta = 1/\sqrt{2}$), each weak state has equal amounts of the two mass states.

To see how oscillations can occur, we must describe the time evolution of a free neutrino. Consider a muon neutrino produced by the weak force at $t = 0$. It is a linear combination of two mass states, or matter waves, that are, by the convention in Equation (7) exactly 180 degrees out of phase with

one another. In quantum mechanics, the time evolution of a state is determined by its energy, and the energies of the mass states are simply given by

$$E_k = \sqrt{p^2c^2 + m_k^2c^4} \;, \tag{8}$$

where p is the momentum of the neutrinos and m_k ($k = 1, 2$) is the mass of the states ν_1 and ν_2, respectively. Note that, if the particle is at rest, this is just the famous energy relation of Einstein's special relativity, $E = mc^2$. In quantum mechanics, the time evolution of each mass component ν_k is obtained by multiplying that component by the phase factor $\exp[-i(E_k/\hbar)t]$, and thus the time evolution of the muon neutrino is given by

$$|\nu_\mu(t)\rangle = -\sin\theta \exp[-i(E_1/\hbar)t]|\nu_1\rangle$$
$$+\cos\theta \exp[-i(E_2/\hbar)t]|\nu_2\rangle \tag{9}$$

as discussed in the box "Derivation of Neutrino Oscillations" on the next page. Because the two states $|\nu_1\rangle$ and $|\nu_2\rangle$ have different masses, they also have different energies (E_1 is not equal to E_2), and the two components evolve with different phases.

Figure 14(b) plots the wavelike behavior of each of the mass components (red and yellow) and shows how the relative phase of the two components varies periodically in time. At $t = 0$, the two components add up to a pure muon neutrino (a pure weak state), and their relative phase is π. As their relative phase advances in time, the mass components add up to some linear combination of a muon neutrino $|\nu_\mu\rangle$ and an electron neutrino $|\nu_e\rangle$, and when the relative phase has advanced by 2π, the components add back up to a muon neutrino. The relative phase oscillates with a definite period, or wavelength, that depends on the difference in the energies of the two mass components, or equivalently, the squared mass differences, $\Delta m^2 = m_1^2 - m_2^2$.

In quantum mechanics, observations pick out the particle rather than the wave aspects of matter, and in the case of neutrinos, they pick out the weak-

interaction properties as opposed to the free-propagation characteristics of mass and momentum. So, in an individual measurement of an event, there are only two possibilities: to detect the muon neutrino or the electron neutrino, but not some linear combination. Thus, what is relevant for an experiment is the probability that the muon neutrino remains a muon neutrino at a distance x from its origin, $P(\nu_\mu \to \nu_\mu)$, or the probability that the muon neutrino has transformed into an electron neutrino, $P(\nu_\mu \to \nu_e)$. The box "Derivation of Neutrino Oscillations" on the next page shows how to calculate these probabilities from the time-evolved state. The results are

$$P(\nu_\mu \to \nu_\mu) = 1 - \sin^2 2\theta \sin^2(\frac{\pi x}{\lambda_{osc}}) \tag{10}$$

and

$$P(\nu_\mu \to \nu_e) = \sin^2 2\theta \sin^2(\frac{\pi x}{\lambda_{osc}}) \;, \tag{11}$$

where θ is the mixing angle defined above, x is measured in meters, and λ_{osc} is the oscillation length given in meters. The oscillation length (the distance between two probability maxima or two probability minima) varies with the energy of the neutrino E_ν (in million electron volts), and it also depends on the squared mass difference (in electron volts squared):

$$\lambda_{osc} = 2.5 E_\nu / \Delta m^2 \;, \tag{12}$$

The two probabilities in Equations (10) and (11) oscillate with distance x from the source, as shown in Figure 14(c).

To summarize, a muon neutrino produced at $t = 0$ travels through space at almost the speed of light c. As time passes, the probability of finding the muon neutrino $P(\nu_\mu \to \nu_\mu)$ decreases below unity to a minimum value of $1 - \sin^2 2\theta$ and then increases back to unity. This variation has a periodicity over a characteristic length $\lambda_{osc} \cong cT$, where T is the period of neutrino oscillation. The oscillation length varies inversely with Δm^2. The probability of finding an electron neutrino in place

Derivation of Neutrino Oscillations

Some simple algebra can show how neutrino oscillation effects depend on the mass difference of the neutrino mass eigenstates. Consider the simplified case of just two neutrino flavors. We express the quantum mechanical wave function for a muon neutrino produced at $t = 0$ as a mixture of the mass eigenstates $|\nu_1\rangle$ and $|\nu_2\rangle$ with masses m_1 and m_2, respectively.

$$|\nu_\mu(0)\rangle = |\nu_\mu\rangle = -\sin\theta\,|\nu_1\rangle + \cos\theta\,|\nu_2\rangle \ ,$$

where an electron neutrino is given by $|\nu_e\rangle = \cos\theta\,|\nu_1\rangle + \sin\theta\,|\nu_2\rangle$ and the angle θ characterizes the extent of mixing of the mass eigenstates in the weak-interaction eigenstates. It is called the mixing angle. (For more than two flavors, there are more mixing angles as well as charge-conjugation and parity, CP, violating phases.) At a later time t, the wave function is

$$|\nu_\mu(t)\rangle = -\sin\theta\,\exp(-iE_1 t)|\nu_1\rangle + \cos\theta\,\exp(-iE_2 t)|\nu_2\rangle \ ,$$

where the mass eigenstates propagate as free particles and E_1 and E_2 are the energies of those states $|\nu_1\rangle$ and $|\nu_2\rangle$, respectively. (We are working in units for which $\hbar = c = 1$.) For relativistic neutrinos ($E_\nu \gg m$), we can approximate E_1 and E_2 by

$$E_k = (p^2 + m_k^2)^{1/2} \cong p + m_k^2/2p \ ,$$

where we are assuming that the two mass states have the same momentum. After substituting these energies, the wave function at time t becomes

$$|\nu_\mu(t)\rangle = \exp[-it\,(p + m_1^2/2E_\nu)]\,[-\sin\theta\,|\nu_1\rangle + \cos\theta\,|\nu_2\rangle\,\exp(i\Delta m^2 t/2E_\nu)] \ ,$$

where $\Delta m^2 = m_1^2 - m_2^2$ and $E_\nu \cong p$. Since these neutrinos are traveling almost at the speed of light, we can replace t by $x/c = x$, where x is the distance from the source of muon neutrinos. Let us now calculate $P(\nu_\mu \to \nu_e)$, which is defined as the probability of observing a ν_e at x, given that a ν_μ was produced at the origin $x = 0$. The probability is the absolute square of the amplitude $\langle \nu_e | \nu_\mu(t)\rangle$. Using the orthonormality relation $\langle \nu_i | \nu_j \rangle = \delta_{ij}$, we can compute the probability

$$P(\nu_\mu \to \nu_e) = |\cos\theta\,\sin\theta\,(1 - \exp(i\Delta m^2 t/2E_\nu)|^2$$

$$= \sin^2 2\theta\,\sin^2\left(\Delta m^2 x/4E_\nu\right)$$

$$= \sin^2 2\theta\,\sin^2\left(\frac{1.27\Delta m^2 x}{E_\nu}\right) \ ,$$

where Δm^2 is measured in electron volts squared, x is in meters, and E_ν is in million electron volts, and the factor of 1.27 derives from working in these units. $P(\nu_\mu \to \nu_\mu)$ is the probability of observing a ν_μ at x, given that a ν_μ is produced at $x = 0$. This probability can be computed explicitly, or by the conservation of probability, it is

$$P(\nu_\mu \to \nu_\mu) = 1 - \sin^2 2\theta\,\sin^2\left(\frac{1.27\Delta m^2 x}{E_\nu}\right) \ .$$

It is often useful to define an oscillation length, λ_{osc} for these probabilities, which, as shown in Figure 14(c), equals the distance between the maxima (or the minima). Note that the spatial period of $\sin^2 x$ is one-half that of $\sin x$, and so the oscillation length is defined by the following equation:

$$P(\nu_\mu \to \nu_e) = \sin^2 2\theta\,\sin^2\left(\frac{1.27\Delta m^2 x}{E_\nu}\right) = \sin^2 2\theta\,\sin^2\left(\frac{\pi x}{\lambda_{osc}}\right) \ ,$$

where

$$\lambda_{osc} = \frac{\pi E_\nu}{1.27\Delta m^2} \approx \frac{2.5 E_\nu}{\Delta m^2} \ .$$
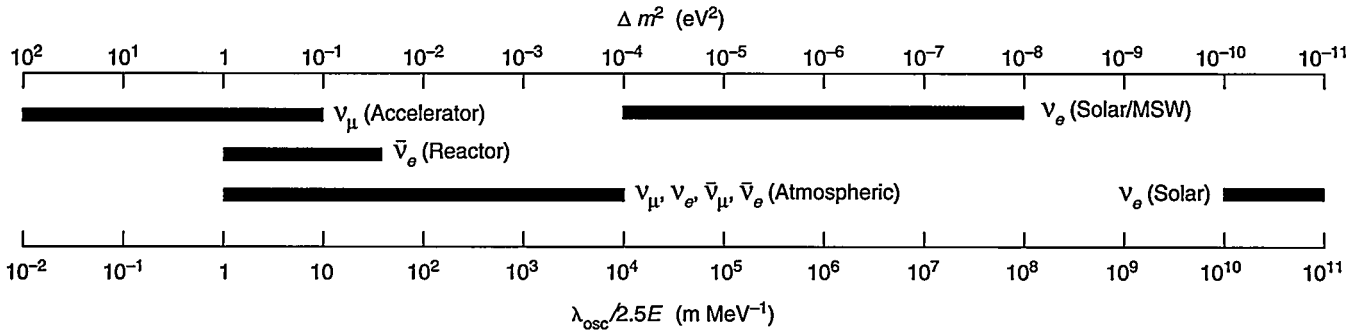
Figure 15. Accessible Ranges of Δm^2

Neutrino energies are specific to the source, and source-to-detector distances also vary with the source. The ratio of these two variables determines the range of values for Δm^2 that neutrino oscillation experiments can measure using each source. These ranges are labeled with the source and the neutrinos produced by that source. Two ranges are given for solar-neutrino experiments. One assumes that the MSW effect enhances oscillations, in which case, the range of Δm^2 is determined in part by the electron density of matter in the Sun. The other assumes no matter enhancement.

of the muon neutrino $P(\nu_\mu \to \nu_e)$ also oscillates as a function of distance from the source and has the same wavelength λ_{osc}. That probability has a maximum value of $\sin^2 2\theta$. These formulas show explicitly that, if neutrinos oscillate between family types, neutrinos must have nonzero masses and the neutrino weak states are not states of definite mass but rather mixtures of mass states.

Although we have restricted the analysis to mixing between two families, there is every reason to expect that, if mixing takes place among the leptons, it would occur among all three families and that there would be a mixing matrix for the leptons analogous to the CKM matrix for the quarks. The three-flavor mixing problem is more difficult, but it boils down to carrying out the analysis, which is a technical problem.

Interpreting Oscillation Experiments

Most extensions of the Standard Model tell us to expect mixing among leptons in analogy with mixing among quarks. But so far, those theories make no quantitative predictions on masses and mixing angles. Thus, neutrino oscillation experiments have a twofold purpose: first to establish convincing evidence for oscillations and then to

make quantitative determinations of the neutrino masses and mixing angles.

Among the quarks, the amount of mixing is small and occurs primarily between the first two families. It is natural to assume the same should hold for the leptons, although theory provides no such restriction. Consequently, neutrino oscillation experiments have traditionally been interpreted in the two-family context. Applying the two-family formalism to each experiment allows one to derive a range of possible values for Δm^2 and a range for $\sin^2 2\theta$, where θ is the mixing angle between the two families. Input to the interpretation includes the neutrino energies in a particular experiment, the distance from source to detector, the expected neutrino flux, and the measured flux or probability. In a disappearance experiment, one measures $P(\nu_i \to \nu_i)$, the probability of finding the original neutrino flavor ν_i, where $i = e, \mu, \tau$. In an appearance experiment, one measures the probability of finding a flavor different from the original $P(\nu_i \to \nu_j)$, where $i \neq j$.

The only definite constraints on neutrino masses are the following upper limits: $\nu_e < 10$ electron volts (eV), derived from tritium beta decay, $\nu_\mu < 170$ kilo-electron-volts (keV), derived from pion decay, and $\nu_\tau < 24$ MeV, derived from tau decay. So, the field is wide open for exploration.

Figure 15 shows the regions of Δm^2 (and its inverse, $\lambda_{osc}/2.5E_\nu$) that can be probed with the neutrinos from reactors, accelerators, the upper atmosphere, and the Sun. Variations in neutrino energies and source-to-detector distances make each type of experiment sensitive to a different range of values. The largest mass difference accessed by solar-neutrino experiments (assuming the MSW effect) is below the lowest value accessed by other experiments. Given the electron densities in the Sun and the energies of solar neutrinos (1 to 10 MeV), MSW enhancement can take place only for very small values of Δm^2 from 10^{-4} to 10^{-9} eV^2, with a favored value on the order of 10^{-5} eV^2. So, if neutrino oscillation is the explanation behind the solar- and atmospheric-neutrino deficits as well as the LSND appearance measurements, the two-family analysis must be extended to three families. All the data supporting neutrino oscillations are reviewed in the article "The Evidence for Oscillations" on page 116.

We will give one simple example of a model that fits the oscillation data consistently, but there are many other such models with no way to choose among them. This simple model assumes the traditional mass hierarchy, $m_1 < m_2 < m_3$. But the first two masses, m_1 and m_2, are assumed to be very nearly identical and therefore almost

equally distant from the third mass m_3:

$$\Delta m_{12}^2 \approx 10^{-5} \ (eV)^2$$
$$\Delta m_{23}^2 \approx 0.3 \ (eV)^2 \qquad (12)$$
$$\Delta m_{13}^2 \approx 0.3 \ (eV)^2 \ .$$

Thus, there are two distinct oscillation lengths differing by 4 orders of magnitude. Since the upper limit on the electron neutrino mass is 10 eV, all neutrinos in this model would be very light and nearly degenerate in mass. The model also assumes that the mixing angle between the second and third families is close to the maximum value of $\pi/4$, whereas the mixing angles between the first two families and the first and third families are quite small. (Note that this mixing pattern is quite unlike the CKM matrix for the quarks, in which the mixing angle for the second and third families is very small.)

Both LSND and solar-neutrino experiments measure the oscillation of the muon neutrino to the electron neutrino $P(\nu_\mu \rightarrow \nu_e)$, or vice versa, so one might naively assume that both measure Δm_{12}^2, the difference between neutrino masses in the first and second families. But the LSND results for Δm^2 differ by at least 4 orders of magnitude from the solar results. How can the two be reconciled? The resolution comes about because mixing occurs among the three families. Then, three oscillatory terms can contribute to $P(\nu_\mu \rightarrow \nu_e)$, one with an oscillation length determined by Δm_{12}^2 and two others with oscillation lengths determined by Δm_{13}^2 and Δm_{23}^2, respectively.

The source-to-detector distance (30 meters), combined with the neutrino energies, makes LSND sensitive to the two terms whose oscillations are determined by Δm_{13}^2 and Δm_{23}^2. Those $\nu_\mu \leftrightarrow \nu_e$ oscillations take place indirectly through ν_τ. These "indirect oscillations" do not contribute to the solar-neutrino deficit because the wavelengths determined by Δm_{13}^2 and Δm_{23}^2 are too large. The resulting oscillation cannot be amplified by the MSW effect. Instead, the solar electron neutrinos

oscillate directly to muon neutrinos with no involvement of tau neutrinos. Although the intrinsic amplitude for this process is very small (small mixing angle θ_{12}), the amplitude is enhanced by the MSW effect. Solar experiments are thus a measure of Δm_{12}^2. That mass difference is quite small, corresponding to a long oscillation length, and it therefore does not contribute to the LSND results.

Finally, atmospheric-neutrino oscillations are explained by muon neutrinos oscillating into tau neutrinos, a pathway dominated by Δm_{23}^2 and a large mixing angle. This consistent set of mixing angles and mass differences for the neutrinos was outlined by Cardall and Fuller (1996). The specifics of their solution are not as important as the fact that neutrino oscillations could explain the results coming from solar, atmospheric, and accelerator neutrino experiments.

What If Neutrinos Have Mass?

As data accumulate and the evidence for oscillations grows stronger, it is appropriate to examine the implications of lepton mixing. In terms of weak-interaction physics, individual-lepton-family number would no longer be strictly conserved, and the forbidden processes listed earlier could occur. Figure 16 illustrates how the oscillation of a muon neutrino into an electron neutrino would facilitate the process $\mu^- \rightarrow e^- + \gamma$. Unfortunately, the predicted rate for the process in Figure 16, in which the mixing occurs through neutrino interactions with the Higgs background, is far below the limit of detectability, about 10^{-40} times the rate of ordinary muon decay.[7] More generally, lepton-family mixing through interaction with the Higgs bosons would parallel the mixing seen among quarks and lend further support to the idea presented in the Grand Unified Theories that quarks and leptons are close relatives.

In terms of the neutrino itself, oscillations would imply nonzero neutrino

masses, and therefore the basic description of the neutrino would have to be altered. The neutrino might be a Dirac particle and parallel the Dirac electron in having four independent states—right-handed and left-handed particle states, ν_R and ν_L, and right-handed and left-handed antiparticle states, $\bar{\nu}_R$ and $\bar{\nu}_L$. To complete this set of four, two new neutrino states would have to be added to the Standard Model: the right-handed neutrino ν_R and the left-handed antineutrino $\bar{\nu}_L$. The new states would be "sterile" in the sense that they would not interact through the weak force (or any other known force except gravity), and they would be included in the theory only as necessary ingredients to give the Dirac neutrino a mass.

Those sterile neutrino states, however, could differ in mass from the ordinary neutrino states that couple to the W, in which case the ordinary neutrinos could oscillate into those sterile, noninteracting forms. That possibility could have an impact in various astrophysical and cosmological contexts, and conversely, cosmological arguments would place limits on the existence of such sterile neutrinos.

On the other hand, the neutrino might be a Majorana particle, which, by definition, has just two particle states. The two observed states (left-handed neutrino ν_L and right-handed antineutrino $\bar{\nu}_R$) would be the full set. But they would have a new property that would make them freaks in the pantheon of elementary spin-1/2 particles—they could transform into each other and, in effect, would be their own antiparticles. As a result, the weak force could transform an electron into a left-handed electron neutrino, as usual, but then that left-handed neutrino could later appear as *a right-handed antineutrino* and interact through the weak force to

---

[7] Perhaps new forces, such as those expected in supersymmetric theories, also cause transitions between families and contribute to the process $\mu \rightarrow e + \gamma$. For a discussion of how new forces could contribute to muon decay, see the article "The Nature of Muon Decay and Physics beyond the Standard Model" on page 128.

Initial state   Final state

Gamma γ

W^- W^-

μ^- ν_μ ν_e e^-
Muon   Electron

Interaction with
Higgs bosons
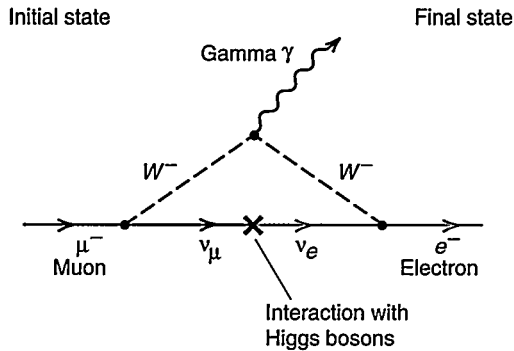
Figure 16. Example of Lepton-Family Mixing

If neutrinos have mass and lepton-family number is not conserved, a muon neutrino ν_μ emitted at the first weak-interaction vertex could become an electron neutrino ν_e through interaction with the Higgs background and be transmuted into an electron e^- at the second vertex. Thus, the reaction $\mu^- \to e^- + \gamma$ could proceed if mixing occurred across lepton families.

Inverse Beta Decay

Proton   Neutron

Electron
e^-

W^+

ν_L $\bar{\nu}_R$

W^+ e^+
            Positron

Proton   Neutron   Majorana mass term
Electron Capture   (interaction with Higgs
background flips handedness,
changing particle to antiparticle)
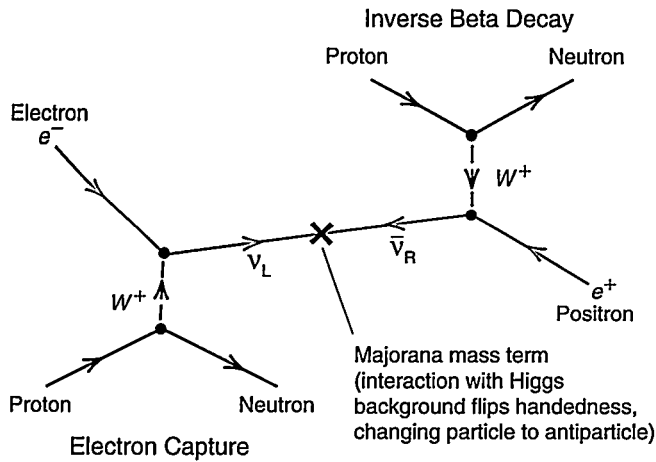
Figure 17. Example of Lepton-Number Nonconservation

If neutrinos are Majorana particles, a left-handed neutrino emitted in electron capture could become a right-handed antineutrino and create a positron through inverse beta decay. Such a process would change lepton number by two units. Notice that the left-handed neutrino flips its handedness through interaction with the Higgs background. This example of lepton-number violation should be compared with the example of lepton-number conservation in Figure 11.

become a positron. Such particle-antiparticle transitions would violate the law of total-lepton-number conservation as well as individual-family-number conservation (see Figure 17). They would also make possible a new type of beta decay known as neutrinoless double beta decay. Unfortunately, that process may be the only measurable sign that the neutrino is a Majorana rather than a Dirac particle.

If it is so hard to tell the types of neutrinos apart at low energies, why

should one care one way or the other? First, if neutrinos were Majorana particles, there would be no new low-mass neutrino states, and the number of mass and mixing-angle parameters in the theory would be highly restricted. This, in turn, would put strong constraints on any theoretical fit to the neutrino oscillation data. Second, the difference between Majorana and Dirac neutrinos is directly related to how neutrinos acquire mass. Grand Unified Theories and most

other extensions to the Standard Model suggest that the familiar neutrinos are Majorana particles and that they have very heavy relatives that reduce their masses through, what is sometimes called, the seesaw mechanism (explained later in this article).

Handedness versus Helicity. To elaborate further on these issues, we must consider the esoteric concept of handedness, a two-valued quantity related in a nontrivial way to helicity. Helicity and handedness are identical for massless particles and almost identical for massive particles, those traveling close to the speed of light. But the concept of handedness is crucial because (1) the weak force of the W distinguishes between different values of handedness and (2) the origin of particle masses and the fundamental differences between Dirac and Majorana neutrinos also involve the concept of handedness.

Figure 18 displays the helicity and handedness states of the electron and the massless electron neutrino as they appear in the Standard Model. Helicity is easy to describe. It is the polarization, or projection, of a particle's intrinsic spin along its direction of motion. There are two such states: spin along the direction of motion (right helicity, or motion like a right-handed corkscrew) and spin opposite to the direction of motion (left helicity, or motion like a left-handed corkscrew). A particle can be produced in a state of definite helicity, and because angular momentum is conserved, that state can be measured directly. The problem is that, for particles with mass, helicity is not a relativistically invariant quantity: As shown in the cartoon on page 57, if neutrinos have mass, then their helicity can change with the reference frame.

In contrast, handedness (also called chirality), although harder to define without using the Dirac equation for spin-1/2 particles, provides a relativistically invariant description of a particle's spin states. There are two independent handedness states for

spin-1/2 particles—left and right. A purely left-handed state has $N_x = L$, a purely right-handed state has $N_x = R$, and like lepton number and electric charge, a particle's handedness is independent of the reference frame from which it is viewed. Further, a particle, massless or massive, can be decomposed into two independent components, left-handed and right-handed, and this decomposition does not change with the reference frame.

The confusing thing about handedness is that it is not a constant of the motion; a spin-1/2 particle traveling through space can change its handedness without changing its helicity. Nevertheless, because it is relativistically invariant, handedness is an essential quantity for describing the properties of the weak force and the origin of particle masses, as well as particle properties. For example, when we say that interactions involving the W are

Figure 18. Helicity and Handedness

(a) The Standard Model. The four helicity and the four handedness (chirality) states of the electron are illustrated here. The strikers between these states indicate that each handedness state can be written as a linear combination of helicity states. The neutrino has only two states, and because it is massless, its helicity is identical to its handedness. Recall that the spin can be represented by a pseudovector (red arrow) and that its direction relative to the momentum determines helicity.

(b) Neutrinos with mass. The states of the Dirac neutrino versus those of the Majorana neutrino are shown. Like the electron, the Dirac neutrino has four helicity and four chirality states. The Majorana neutrino has only two handedness and two helicity states. Further, no clear distinction exists between particle and antiparticle.

(c) Effects of mass terms on particle states. Mass terms always flip the handedness of a particle. The Dirac mass term conserves lepton number or particle number, whereas a Majorana mass term changes particle into antiparticle as it changes the handedness. A Majorana mass term is allowed only for neutral particles. The mass term(s) of the neutrino could be Majorana, Dirac, or both combined.

left-handed, we mean that they pick out the left-handed components of particles and the right-handed components of antiparticles. The correct Standard Model description of the weak isospin doublets, therefore, includes handedness labels. In the first family, the left-handed electron e_L and the left-handed electron neutrino ν_{eL} form a weak isospin doublet. The right-handed electron e_R exists but does not interact with the W, and it is therefore called a weak isospin singlet (that is, it is a scalar quantity under the weak isospin symmetry, and its weak isotopic charge is zero). The right-handed electron does couple electromagnetically to the photon and weakly to the Z^0.

The right-handed (right-helicity) neutrino ν_{eR} does not exist in the Standard Model, but if included, it would be a weak isospin singlet and thus sterile in the sense already described. Similarly, because the weak force picks out the right-handed components of antiparticles, the right-handed positron $\bar{e}_R$ and the right-handed electron antineutrino $\bar{\nu}_{eR}$ form a weak isospin doublet. The left-handed positron $\bar{e}_L$ is a weak isospin singlet and has no weak isotopic charge. Table III lists the weak isospin doublets, the weak isospin singlets, and their charges for the first family of weak states in the Standard Model. The charges of the Higgs doublet h^0 and h^+ are also listed. (The Higgs doublet is discussed in the sidebar "Neutrino Masses" on page 64.)

Handedness is also a crucial concept for the discussion of neutrinos masses. It can be shown mathematically that any interaction or mechanism that gives spin-1/2 particles a nonzero rest mass must connect particles of different handedness, that is, the interaction must annihilate a particle of one handedness and create a particle of the opposite handedness. Thus, a particle with mass may switch between right- and left-handed states as it travels through space, changing its handedness by two units, whereas a massless particle undergoes no such transformation and maintains both its handedness and helicity.
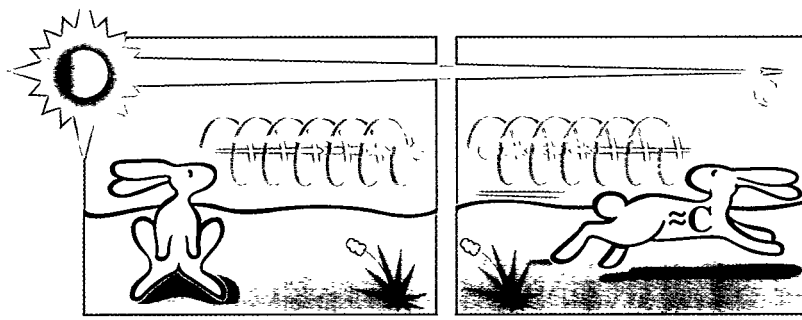
Looks like a left-handed corkscrew.   No—like a right-handed corkscrew!

Table III. First-Family Weak States and Electroweak Charges in the Standard Model

Particle Number N	Handedness N_x	Particle States	Weak Isotopic Charge I_3^W	Weak Hypercharge Y^W	Electric Charge Q	
						QUARKS
$+1$	L	$\begin{pmatrix} u_L \\ d_L \end{pmatrix}$ u_L	$+1/2$	$+1/3$	$+2/3$	
$+1$	L	d_L	$-1/2$	$+1/3$	$-1/3$	Weak isospin doublets
-1	R	$\begin{pmatrix} \bar{u}_R \\ \bar{d}_R \end{pmatrix}$ $\bar{u}_R$	$-1/2$	$-1/3$	$-2/3$	
-1	R	$\bar{d}_R$	$+1/2$	$-1/3$	$+1/3$	
$+1$	R	u_R	0	$+4/3$	$+2/3$	
-1	L	$\bar{u}_L$	0	$-4/3$	$-2/3$	Weak isospin singlets
$+1$	R	d_R	0	$-2/3$	$-1/3$	
-1	L	$\bar{d}_L$	0	$+2/3$	$+1/3$	
						LEPTONS
$+1$	L	$\begin{pmatrix} e_L \\ \nu_L \end{pmatrix}$ e_L	$-1/2$	-1	-1	
$+1$	L	ν_L	$+1/2$	-1	0	Weak isospin doublets
-1	R	$\begin{pmatrix} \bar{e}_R \\ \bar{\nu}_R \end{pmatrix}$ $\bar{e}_R$	$+1/2$	$+1$	$+1$	
-1	R	$\bar{\nu}_R$	$-1/2$	$+1$	0	
$+1$	R	e_R	0	-2	-1	
-1	L	$\bar{e}_L$	0	$+2$	$+1$	Weak isospin singlets
$+1$	R	ν_R	0	0	0	
-1	L	$\bar{\nu}_L$	0	0	0	
						HIGGS BOSONS
0	0	$\begin{pmatrix} h^+ \\ h^0 \end{pmatrix}$ h^+	$+1/2$	$+1$	$+1$	Weak isospin doublet
0	0	h^0	$-1/2$	$+1$	0	

R = Right-handed
L = Left-handed

$$Q = I_3^W + \frac{Y^W}{2}$$

Although the mathematical definition of handedness is beyond this discussion, we can get a more concrete idea by seeing the purely left-handed and purely right-handed states of, say, the electron written in terms of helicity states $|e_\lambda\rangle$:

$$|e_L\rangle \propto |e_{-1/2}\rangle + m/E\ |e_{1/2}\rangle \ ,$$
$$|e_R\rangle \propto |e_{1/2}\rangle - m/E\ |e_{-1/2}\rangle \ ,$$

(13)

where m is the mass of the particle, E is its energy, and $\lambda = s \cdot p/|p|$ is the helicity (with right- and left-handed projections of 1/2 and $-1/2$, respectively). These formulas show that, if a particle is massless ($m = 0$), helicity and handedness are identical. And, if a left-handed particle is relativistic, or traveling at nearly the speed of light (m very much less than E), it is mostly in a state of left helicity; similarly, a right-handed particle traveling at relativistic speeds is mostly in a state of right helicity. Handedness and helicity are very much related, yet the two have quite different properties.

To see a tangible effect of those differences, consider the decay of the positively charged pion. This particle decays through the weak force into a lepton and an antilepton, either a positron and an electron neutrino ($\pi^+ \to e^+ + \nu_e$) or an antimuon and a muon neutrino ($\pi^+ \to \mu^+ + \nu_\mu$). The decay into a positron yields more kinetic energy because the positron is lighter than the antimuon; so, if all else were equal, that decay would be more probable than the decay into a antimuon. Yet the opposite is true precisely because handedness and helicity are different. The pion has an intrinsic spin of zero, so for the decay of a pion at rest to conserve both angular and linear momentum, the spins and momenta of the two leptons must point in opposite directions (see Figure 19). In other words, the two leptons must be in the same helicity state. But the decay process occurs through the left-handed weak force and therefore produces a right-

Initial state — Final state

$$\pi^+ \longrightarrow \mu^+ + \nu_\mu$$

$J_z = s_z = -\frac{1}{2}$
The massless, left-handed neutrino is in a state of left helicity.

$J_z = s_z = \frac{1}{2}$
The right-handed antimuon must be produced in a state of left helicity.

Angular momentum $\quad J_z = 0 \longrightarrow J_z = -\frac{1}{2} + \frac{1}{2} = 0$
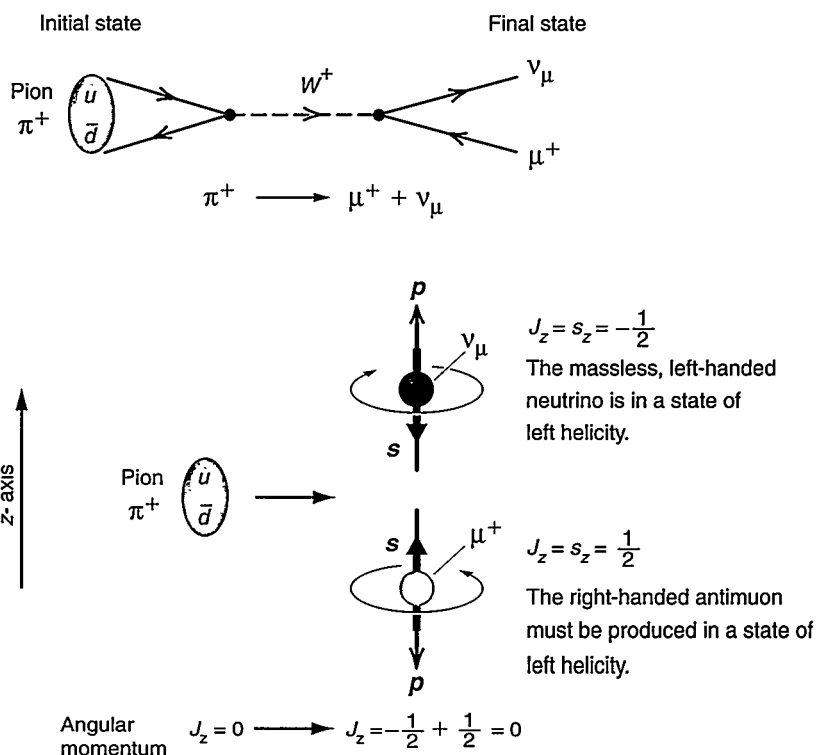
Figure 19. Pion Decay and Helicity versus Handedness

A π^+ has spin zero ($s = 0$). Through the weak force, it decays at rest into a μ^+ and a ν_μ. To conserve total momentum ($p = 0$) and total angular momentum ($J = 0$), these two particles must be emitted with equal and opposite momentum (black arrows), and their spins (red arrows) must point in opposite directions. The neutrino is emitted as a left-helicity particle because it is nearly massless. Thus, the μ^+ must also be in a state of left helicity. But the weak force produces only right-handed antiparticles. The decay shown here proceeds because a *right-handed* antimuon has a small component in the state of *left helicity*.

handed charged antilepton (antimuon or positron) and a left-handed lepton (muon neutrino or electron neutrino, respectively). The left-handed neutrino is massless, or nearly so, and from the formulas above, it must be in a state of left helicity. Therefore, only the fraction of the right-handed charged antilepton that is in the state of left helicity can take part in the decay. Being proportional to m/E, the left-helicity fraction is much larger for the antimuon than for the positron. Since the decay rate is proportional to the square of that fraction, the pion decays into an antimuon about 10^4 times more frequently than into a positron!

Dirac versus Majorana Neutrinos— Adding Neutrino Masses to the Standard Model. Knowing how handedness and helicity differ for particles with mass, we can return to the question of Majorana versus Dirac neutrinos. Were the neutrino truly massless, there would be no way to tell whether it is a Dirac or Majorana particle. Either way, there would be two neutrino states: ν_L and $\bar\nu_R$. Each would travel at the speed of light, and each would maintain its handedness (and its helicity) independent of the observer's reference frame. Either way, the weak isospin doublets would be (ν_{eL}, e_L) and $(\bar\nu_{eR}, \bar{e}_R)$ as defined in Table III, and the members within each weak doublet would transform into

each other under the weak force.

The difference in the properties of a Dirac versus Majorana neutrino has to do with the way in which the neutrino acquires its mass. We already said that, whatever mechanism gives a spin-1/2 particle its mass, it must change that particle's handedness by two units, from left to right or vice versa. Figure 20(a) illustrates how the electron, a Dirac particle with four states, acquires its mass in the Standard Model. The interaction is between the Higgs background (this is the Higgs mechanism that gives mass to all particles in the Standard Model) and the electron. Called a Dirac mass term, this interaction annihilates the state e_L and creates the state e_R, or it annihilates $\bar{e}_R$ and creates $\bar{e}_L$. In each case, the mass term changes the handedness by two units, as required for any mass term. But it preserves the particle's electric charge and lepton number because a particle state remains a particle state and an antiparticle state remains an antiparticle state.

Note that a mass term for the electron that changed e_L into $\bar{e}_R$ is not an allowed mechanism for giving electrons their mass, even though it changes handedness by two units. It would change a negatively charged electron into a positively charged antielectron (positron), violating electric-charge conservation. But electric charge is known to be conserved. Such a term would also violate total-lepton-number conservation and electron-family-number conservation.

Now consider the neutrino. Since the neutrino has no electric charge, it has several possible mass terms. The diagrams in Figure 20(b) illustrate the interactions that might be added to the Standard Model to give mass to these neutral particles.[8]

The first is a Majorana mass term

that again involves the Higgs background, but it acts between the two neutrino states already available in the Standard Model, changing ν_L into $\bar{\nu}_R$. The Majorana mass term changes both handedness and lepton number by two units. Since it changes a neutrino state into an antineutrino state, the distinction between particle and antiparticle becomes blurred. The neutrino becomes a Majorana particle, or its own antiparticle. This option requires no new neutrino states. The particular term shown in Figure 20(b.1) is called an "effective" theory, good only at low energies because, like Fermi's original theory of beta decay, it gives physically inconsistent answers at high energies.

Figure 20(b.2) pictures the second approach: introducing a Dirac mass term for the neutrino analogous to that shown in Figure 20(a) for the electron. It would change ν_L into ν_R and $\bar{\nu}_R$ into $\bar{\nu}_L$. In other words, it would conserve lepton number. The Dirac mass term requires the introduction of the sterile states ν_R and $\bar{\nu}_L$, and the neutrino becomes a Dirac particle.

The Seesaw Mechanism for Making Neutrino Masses Very Small. The problem with the second approach is that it does not explain why the neutrino masses are so small. In the Standard Model, particle masses are proportional to the strengths of the interactions between the particles and the Higgs bosons (see the box "Family Mixing and the Origin of Mass" on page 72). Thus, the Dirac mass term for, say, the electron neutrino must be multiplied by some very small coupling strength such that the mass of the electron neutrino is at least 50,000 times smaller than the mass of the electron. But the electron and the electron neutrino are part of the same weak doublet, and there seems to be no reason why they should have such enormously different interaction strengths to the Higgs bosons.

In 1979, without introducing an arbitrarily small coupling strength to the

(a) Electron Mass

$$e_L \longrightarrow \times \longrightarrow e_R$$
$<h^0>$ (Higgs background)
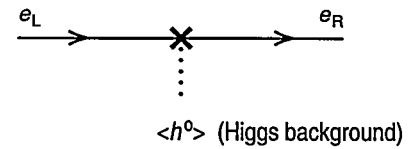
(b) Neutrino Mass

1. Effective Majorana mass term

$<h^0>$ $<h^0>$

$$\nu_L \longrightarrow \times \longleftarrow \bar{\nu}_R$$

2. Dirac mass term

$$\nu_L \longrightarrow \times \longrightarrow \nu_R$$
$<h^0>$
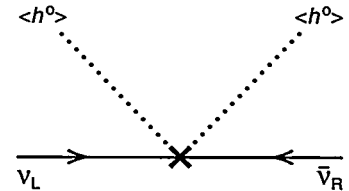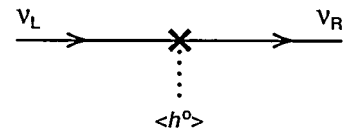
3. Seesaw mass term (Dirac plus Majorana)

$<h^0>$ $<h^0>$

ν_R $\bar{\nu}_L$

ν_L M (large Majorana mass) $\bar{\nu}_R$
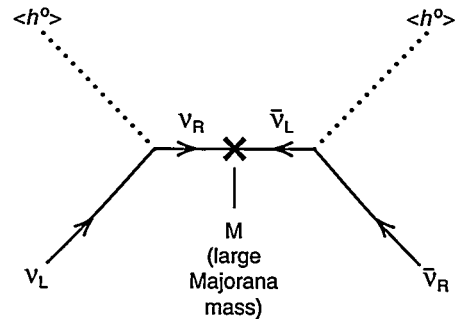
4. Majorana mass term (couples to new boson)

$$\nu_L \longrightarrow \times \longleftarrow \bar{\nu}_R$$
$<\phi^0>$
Nambu-Goldstone boson (from Higgs triplet)
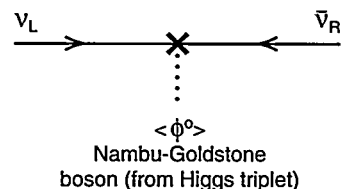
Figure 20. Neutrino Mass Terms
The figures above illustrate the mechanisms for giving the neutrino its mass. In each case, the X represents the effect from the Higgs background. The direction of time is from left to right.

---

[8]These extensions are explained more fully in the sidebar "Neutrino Masses—How to Add Them to the Standard Model" on page 64. They provide the simplest way of including nonzero neutrino masses while preserving the local gauge symmetries.

Higgs bosons, Murray Gell-Mann, Pierre Ramond, and Richard Slansky invented a model that yields very small neutrino masses. As explained in "Neutrino Masses," the two neutrino states ν_R and $\bar{\nu}_L$ that must be added to the theory to form the Dirac mass term could themselves be coupled to form a Majorana mass term. That term could also be added to the theory without violating any symmetry principle.

Further, it could be assumed that the coefficient M of the Majorana mass term is very large. If the theory contains both a Dirac mass term and this Majorana mass term, then the four components of the neutrino would no longer be states of definite mass m determined by the coefficient of the Dirac mass term. Instead, the four components would split into two Majorana neutrinos, each made up of two components. One neutrino would have a very small mass, equal to m^2/M from the mass term in Figure 20(b.3); the second neutrino would have a very large mass, approximately equal to M. The very light Majorana neutrino would mostly be the left-handed neutrino that couples to the W, and the very heavy neutrino would mostly be a right-handed neutrino that does not couple to the W. Similarly, the very light antineutrino would be mostly the original right-handed antineutrino that couples to the W, and the very heavy antineutrino would be mostly a left-handed antineutrino that does not couple to the W.

This so-called seesaw mechanism in which the Dirac mass m is reduced by a factor of m/M through the introduction of a large Majorana mass term has been used in many extensions to the Standard Model to explain why neutrino masses are small. The large Majorana mass M is often associated with some new, weak gauge force that operates at a very high energy (mass) scale dictated by the mass of a new, very heavy gauge boson. The net result of this approach is that the neutrino seen at low energies is predicted to be mostly a Majorana particle!

Figure 20(b.4) shows one last possibility for adding neutrino masses to the Standard Model. No new neutrino components are added to the Standard Model. Instead, the neutrino is postulated to be a two-component Majorana particle that acquires mass by coupling to a new type of Higgs boson, one that has three charge states and is a triplet in a weak isospin space. Thus, introducing a new type of Higgs boson allows neutrino masses to be added. This last possibility has several interesting consequences. Total-lepton-number conservation is not explicitly violated by the addition of a Majorana mass term. Instead, the new Higgs boson is assumed to have a nonzero vacuum value; the resulting Higgs background spontaneously breaks lepton-number conservation and gives a Majorana mass to the neutrino. A consequence of this spontaneous (or vacuum) breaking of the lepton-number symmetry is the existence of a massless scalar particle known as a Nambu-Goldstone boson. This massless boson could be produced in a new form of neutrinoless double beta decay.

Neutrinoless Double Beta Decay. The one process that should be within the limits of detectability and would exhibit the unmistakable mark of a Majorana neutrino is neutrinoless double beta decay. In double beta decay, two neutrons in a nucleus transform into two protons almost simultaneously and bring the nucleus to a stable configuration with an increase in electric charge of $+2$. This process occurs in "even-even" nuclei, those containing even numbers of protons and neutrons. Like single beta decay, double beta decay occurs through the interaction of a nucleus with the W. In the "ordinary" process shown in Figure 21(a), the nucleus emits two electrons and two antineutrinos. Figure 21(b) shows, however, that if the neutrino is a Majorana particle, the same process can occur without the emission of any neutrinos— hence the name of neutrinoless double beta decay. The weak force has not changed its character. Indeed, when

the first neutron transforms into a proton and emits a W, that W produces a right-handed antineutrino and an electron, as usual. Then that *right-handed antineutrino* switches to a *left-handed neutrino* through the interaction that gives the neutrino its Majorana mass. Finally, this left-handed neutrino then interacts with the second W (emitted when the second neutron transforms into a proton), and the left-handed electron neutrino is transformed into a left-handed electron. The neutrino is never seen; it is a virtual particle exchanged between the two Ws that are emitted when the two neutrons change into two protons simultaneously. The net result is that two neutrons in a nucleus turn into two protons and two electrons are emitted. In this process, the total charge is conserved, but the number of leptons has changed from zero to two. Also, because no neutrinos are emitted, the two electrons will always share all the available energy released in the decay, and thus the sum of their energies has a single value, the single spike in Figure 21(a), rather than a spectrum of values as in ordinary double beta decay.

The rate of neutrinoless double beta decay is proportional to an effective mass that is a complicated sum over the three neutrino masses. This sum involves the intrinsic charge-conjugation and parity properties of the neutrinos (CP parities), and the resulting phases multiplying each mass can lead to cancellations such that the effective mass is smaller than any of the individual masses of the neutrinos. At present, the experimental upper limit on the effective mass is about 2 eV.

Finally, if the neutrino acquires mass through the vacuum value of a Higgs triplet, as discussed above, a massless Nambu-Goldstone boson would be emitted along with the two electrons of the neutrinoless double beta decay. The presence of the massless boson would lead to a definite energy spectrum for the emitted electrons that would distinguish this form of double beta decay from either ordinary double beta decay or neutrinoless double beta decay.
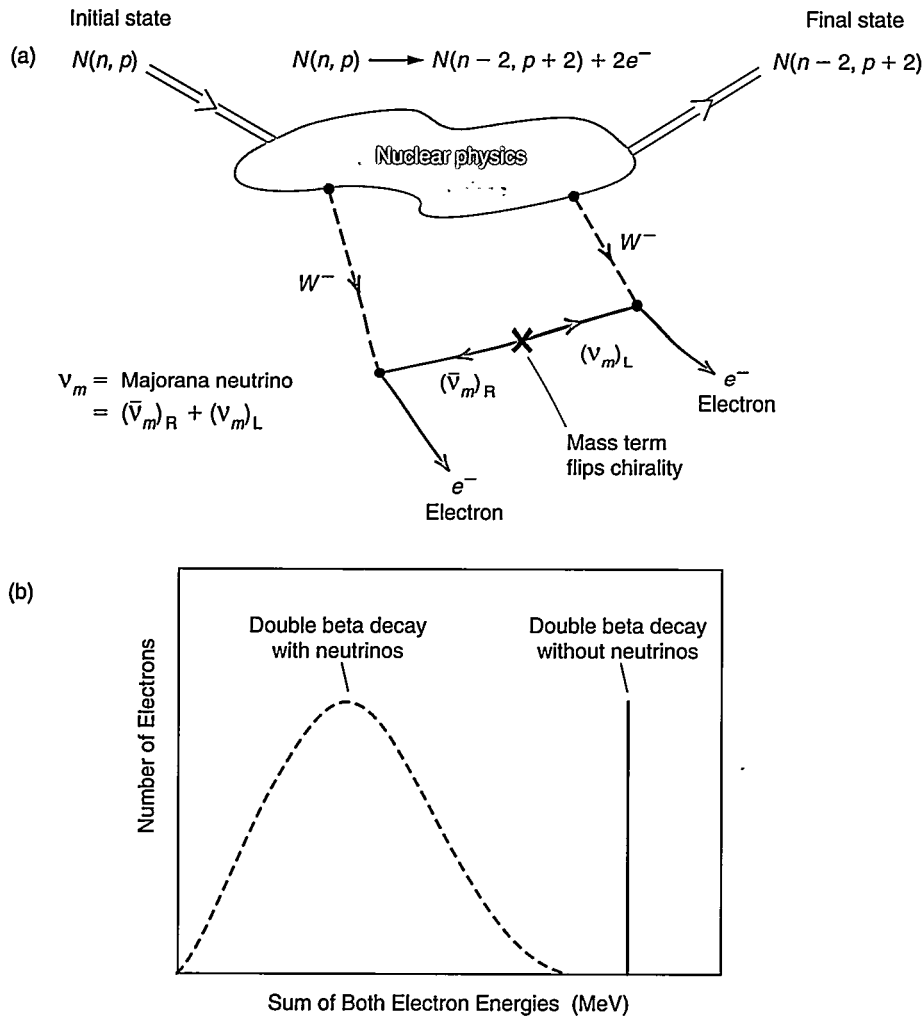
Figure 21. Neutrinoless Double Beta Decay

(a) The exchange of a virtual Majorana neutrino allows double beta decay to occur without the emission of any neutrinos. A right-handed Majorana antineutrino is emitted (along with an electron) from the weak vertex at left. Its handedness flips as it propagates through the interaction with the Higgs background, and the right-handed antineutrino becomes a left-handed Majorana neutrino. In its left-handed form, this particle has the correct handedness to be absorbed at the weak vertex at right and then transformed into an electron. Thus, two electrons are emitted as the nucleus increases its positive electric charge by two units. (b) The spectrum of the total energy carried by two electrons from neutrinoless double beta decay is just a single line because the two electrons always carry off all the available energy (a heavy nucleus absorbs momentum but, essentially, no energy). In contrast, the electrons from ordinary double beta decay share the available energy with the two electron antineutrinos emitted in the decay.

Implications of Neutrino Mass for Astrophysics, Cosmology, and Particle Physics

If neutrino masses and oscillations are real, they can have an impact on astrophysics and cosmology, and, conversely, astrophysics and cosmology will place constraints on the masses of neutrinos and on the number or types of neutrinos. Neutrinos are very weakly coupled to matter. At energies of 1 MeV, a neutrino interacts 10^{20} times less often than a photon. To have any impact at all, they must be present in extraordinary numbers. One such "place" is the universe itself. Neutrinos left over from the Big Bang fill the universe and outnumber protons and neutrons by a billion to one. On average, the universe contains about 300 neutrinos per cubic centimeter, 100 of each of

the three types. If individual neutrino masses are on the order of a few electron volts, their sum would add up to a significant fraction of the mass of the universe—not enough mass to close the universe and have it collapse back on itself (that would require the average mass of the three neutrinos to be 30 eV), but at smaller values, it could have influenced the expansion of matter after the Big Bang and helped produce the superlarge-scale filigree pattern of galaxies and galactic clusters that extends as far as today's telescopes can see. (See the article "Dark Matter and Massive Neutrinos" on page 180.)

Neutrino oscillations, too, may be an important ingredient in making the universe as we know it. For example, the neutrinos we know might oscillate into sterile neutrinos, those which have no weak interactions at all. The presence

of these sterile neutrinos in the cosmic soup could shift the delicate balance of ingredients needed to predict the observed primordial abundances of helium and other light elements up through lithium. As a result, nucleosynthesis calculations place stringent limits on sterile neutrinos, ruling out significant portions in the $\Delta m^2 - \sin^2\theta$ plane for the mixing between ordinary and sterile neutrinos.

Oscillation could also alter the picture of the neutrino as the driver of supernova explosions (see the article "Neutrinos and Supernovae" on page 164). Electron neutrinos, the primary drivers, might be lost or gained from the region that powers the explosion, depending on the oscillation length and, again, on whether sterile neutrinos exist. Neutrino oscillations and the enhancement of those oscillations through interactions with matter

may also be the only way to create the neutron-rich environment that is absolutely required for the synthesis of the elements heavier than iron. And to recap the earlier discussion, oscillation from one neutrino type into another might explain why neutrino physicists have been measuring a shortfall in the ratio of muon neutrinos to electron neutrinos produced by cosmic rays in the upper atmosphere. Matter-enhanced neutrino oscillations in the electron-rich environment of the Sun might explain why physicists observe a shortfall in the flux of electron neutrinos that are produced by thermonuclear fusion processes in the core of the Sun.

Grand Unified Theories. On a more abstract note, the existence of neutrino masses and mixing will extend the close parallel already observed between quarks and leptons and, for that·reason, may well add fuel to the ongoing search for a theory that unifies the strong, weak, and electromagnetic forces. Attempts to explain the pattern of charges and masses of quarks and leptons within a single weak family (columns in Figure 5) lead naturally to an extension of the Standard Model known as the Grand Unified Theories. In these theories, the local gauge symmetries of the weak, strong, and electromagnetic forces are subsumed under a larger local gauge symmetry. That larger symmetry becomes apparent only at the enormous energies and tiny distance scales known as the unification scale. At that scale, the strong, weak, and electromagnetic forces become unified into one force, and the quarks and leptons within a family become members of a particle multiplet that transform into each other under the unified force, just as the members of each weak isospin doublet transform into each other under interaction with the W.

The Grand Unified Theories provide a natural explanation for the different charges (electric, weak, and strong) for particles in a family. In addition, these theories make several successful predictions. Since the strong, weak, and

electromagnetic forces become one at the unification scale, these theories constrain the strengths of the strong, weak, and electromagnetic couplings to be equal at that scale. Thus, one can put the measured values of the weak- and electromagnetic-coupling strengths into the framework of the Grand Unified Theories and predict the strong-coupling strength and the scale of unification.

In the Grand Unified Theories that include a new symmetry, called supersymmetry, the prediction for the strong coupling agrees with all the available data, and the grand unification scale turns out to be on order of 10^{16} GeV. (For comparison, the proton mass ≈ 1 GeV/c^2, and the largest accessible energies at the new accelerator being planned in Europe will be a few times 10^3 GeV.) These supersymmetric theories also predict relations between the masses of the charged quarks and leptons, and these relations are also well satisfied. Neutrino masses are typically not as constrained as charged fermion masses because the neutrino sector contains the possibility of very heavy (as in the seesaw) Majorana masses.

The proton, which is the most stable particle we know, is typically unstable in the Grand Unified Theories and has a lifetime set by the grand unification scale. Supersymmetric Grand Unified Theories predict that the dominant decay mode for the proton is $p \rightarrow K^+ + \bar{\nu}$. The cumulative evidence collected over the next five years at super-Kamiokande will be sensitive to this decay mode with a predicted lifetime on the order of 10^{33} years. Finally, supersymmetric Grand Unified Theories require new particle states, some of which may be observed at high-energy accelerators, specifically, at the new Large Hadron Collider at CERN scheduled for completion in 2002, at the Fermilab Tevatron (an 1,800-GeV machine) following its upgrade in 1999, and at the Hadron Electron Ring Accelerator at DESY (Deutsches Elektronen Synchrotron). These new states can lead to observable lepton family mixing such as

$\mu^- \rightarrow e^- + \gamma$, and they typically provide a candidate for the cold dark matter that may be needed to explain the observed large-scale structures and large-scale motions of the luminous matter.

Superstrings and Conclusions

To tie up our discussion, we will mention superstring theory, one possible truly unified theory that includes not only the electroweak and strong interactions, but also gravity in the sense of a quantum mechanical theory of Einstein's general theory of relativity. Although not yet a full-fledged theory, superstrings have enjoyed significant recent progress. At "low energies" (although they are very high compared with current accelerator energies), superstring theories reduce to models with large gauge symmetries that may unify the electroweak and strong interactions, along with other undiscovered interactions of nature. Although superstrings are insufficiently formulated to predict the parameters of the Grand Unified Theories, the suggestive link between the two makes us pay close attention to the Grand Unified Theories, even in the absence of direct experimental evidence for them. On a less ambitious plane, experimental values for neutrino masses and mixing angles would constrain the parameters of the Grand Unified Theories—particularly when there is a better understanding of the origin of mass and mixing.

No one yet understands why mass states and weak states differ or, even with experimental data on hand, why the pattern of mixing for quarks is as we observe it. Why there should be three repetitive families is likewise mysterious. If we are to develop a unified theory combining the quark and lepton families, we need to solve these unknowns. Neutrino masses and mixings are among the few uncharted realms that may provide important clues to this puzzle. ∎

Further Reading

Bernstein, J. 1969. *The Elusive Neutrino*. United States Atomic Energy Commission–Division of Technical Information. Library of Congress Catalog Card Number: 71-601702.

Cardall, C. Y., and G. M. Fuller. 1996. Can a "Natural" Three-Generation Neutrino Mixing Scheme Satisfy Everything? *Physical Review D* 53: 4421.

Fermi, E. 1933. Tentative Theory of Beta-Radiation. In Strachan, C. 1969. *The Theory of Beta-Decay*. New York: Pergamon Press.

Feynman, R. P., and M. Gell-Mann. 1958. Theory of the Fermi Interaction. *Physical Review* 109 (1): 193.

Halzen, F., and A. D. Martin. 1984. *Quarks & Leptons: An Introductory Course in Modern Particle Physics*. New York: John Wiley & Sons, Inc.

Kayser, B., F. Gibrat-Debu, and F. Perrier. 1989. *The Physics of Massive Neutrinos*. World Scientific Lecture Notes in Physics Vol. 25. Singapore: World Scientific.

Lee, T. D., and C. N. Yang. 1957. Parity Nonconservation and a Two-Component Theory of the Neutrino. *Physical Review* 105 (5): 1671.

Particle Physics. 1984. Los Alamos Science. Summer/Fall. Number 11. Los Alamos National Laboratory document LA-UR-84-2100.

Riordan, M. 1987. *The Hunting of the Quark: A True Story of Modern Physics*. New York: Simon and Schuster/Touchstone.

Strachan, C. 1969. *The Theory of Beta-Decay*. New York: Pergamon Press.

Sutton, C. 1992. *Spaceship Neutrino*. New York: Cambridge University Press.

Weinberg, S. 1995. *The Quantum Theory of Fields*. Volume II. New York: Cambridge University Press.

Acknowledgment

Richard (Dick) Slansky was born in California. Having earned a B.A. from Harvard and a Ph.D. in physics from the University of California at Berkeley (in 1967), Slansky became postdoctoral fellow at Caltech and later assistant professor in the Physics Department at Yale University. In 1974, he joined Los Alamos as a staff member in the newly established Elementary Particles and Field Theory Group in the Theoretical (T) Division. At the Laboratory, his interests turned to unified theories of electroweak and strong interactions—proton decay, neutrino masses, and applying group theory to these unified theories. Slansky was also active in string theory, and he co-authored a two-volume book on the highest weight representations of Kac-Moody algebras. Eventually, he worked on numerous interdisciplinary issues and was appointed adjunct professor of physics at the University of California at Irvine. Slansky then became a Laboratory Fellow and a Fellow of the American Physical Society and the American Association for the Advancement of Science. In 1993, after having been T-Division leader at Los Alamos for four years, Slansky was reappointed as T-Division Director, a position he currently holds.
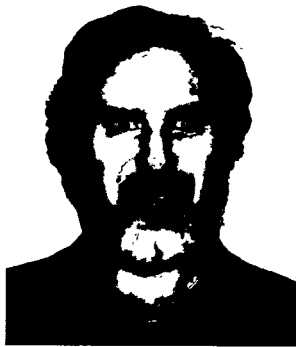
Stuart Raby was born in Bronx, New York, and has been a physics professor at the Ohio State University in Columbus, Ohio, since 1989. He is married to Elaine M. Raby and is a proud parent of Eric and Liat. Raby received his B.Sc. from the University of Rochester in 1969 and an M.Sc. and a Ph.D. from Tel Aviv University in 1973 and 1976, respectively. He was a postdoctoral fellow at Cornell and Stanford Universities as well as a visiting professor at the University of Michigan before joining the Laboratory as a staff member in the Elementary Particles and Field Theory Group, whose leader he became in 1985. His research interests include physics beyond the Standard Model, the origin of fermion masses, flavor violation, and consequences for laboratory experiments, cosmology, and astrophysics. Raby is a leading advocate for supersymmetric Grand Unified Theories of nature. These theories provide the simplest explanation for the experimentally measured unification of the strong, weak, and electromagnetic charges and for the observed spectrum of fermion masses and mixing angles. They also make testable predictions for flavor violation and nucleon decay processes.

T. (Terry) Goldman received a B.Sc. (Hons.) in physics and mathematics in 1968 from the University of Manitoba and an A.M. and a Ph.D. in physics from Harvard in 1969 and 1973, respectively. Goldman was a Woodrow Wilson Fellow from 1968 to 1969 and a National Research Council of Canada postdoctoral fellow at the Stanford Linear Accelerator from 1973 until 1975, when he joined the Laboratory's Theoretical Division as a postdoctoral fellow. In 1978, Goldman became a staff member in the Medium Energy Theory Group at Los Alamos. From 1978 to 1980 he was on leave from the Laboratory as a senior research fellow at the California Institute of Technology, and during the academic year 1982–1983, he was a visiting associate professor at the University of California, Santa Cruz. In 1991, Goldman became group leader in T-5. His research interests are in the areas of electroweak interactions and the quark structure of nuclei. Goldman is a member of the American Physical Society.

Gerry Garvey was born in New York City. He earned a B.S. in physics from Fairfield University and, in 1962, a Ph.D. in physics from Yale University. He then served as a faculty member at Yale and Princeton Universities, becoming a full professor in 1969. After being an Alfred Sloan Foundation Fellow from 1967 to 1969, he spent a year at the Clarendon Laboratory in Great Britain. In 1976, he became Director of the Physics Division at Argonne National Laboratory and in 1979 was named Associate Laboratory Director for Physical Science. From 1979 through 1984, Garvey was a professor at the University of Chicago, returning in 1980 to full-time research as a senior scientist at Argonne. In 1984, he joined Los Alamos as director of LAMPF. In 1990, Garvey stepped down from that position to become a Laboratory Senior Fellow, a position he currently holds. From 1994 through 1996, he served as the Assistant Director for Physical Science and Engineering in the White House Office of Science and Technology. Garvey is an active member of the American Physical Society, having been chairman and councilor for the Division of Nuclear Physics.

Neutrino Masses
How to add them to the Standard Model

Stuart Raby and Richard Slansky

The Standard Model includes a set of particles—the quarks and leptons —and their interactions. The quarks and leptons are spin-1/2 particles, or fermions. They fall into three families that differ only in the masses of the member particles. The origin of those masses is one of the greatest unsolved mysteries of particle physics. The greatest success of the Standard Model is the description of the forces of nature in terms of local symmetries. The three families of quarks and leptons transform identically under these local symmetries, and thus they have identical strong, weak, and electromagnetic interactions.

In the Standard Model, quarks and leptons are assumed to obtain their masses in the same way that the W and Z^0 bosons obtain theirs: through interactions with the mysterious Higgs boson (named the "God Particle" by Leon Lederman). But before we write down some simple formulas that describe the interactions of quarks and leptons with the Higgs boson, let us define some notation.

Defining the Lepton Fields. For every elementary particle, we associate a field residing in space and time. Ripples in these fields describe the motions of these particles. A quantum mechanical description of the fields, which allows one to describe multiparticle systems, makes each field a quantum mechanical operator that can create particles out of the ground state—called the *vacuum*. The act of creating one or more particles in the vacuum is equivalent to describing a system in which one or more ripples in the fabric of the field move through space-time.

Let us now discuss the simple system of one family of leptons. To be specific, we will call the particles in this family the electron and the electron neutrino. The electron field describes four types of ripples (or particles). We label these four types by two quantum charges called fermion number N and handedness, or chirality, N_x. For the electron field, the particle state with fermion number $N = +1$ is the electron, and the particle state with $N = -1$ is the antielectron (or positron). Each of these states comes as right-handed, $N_x = R$, and left-handed, $N_x = L$. Handedness is a Lorentz invariant quantity that is related in a nontrivial way to helicity, the projection of the spin s in the direction of the momentum p. (For a discussion of handedness versus helicity, see "The Oscillating Neutrino" on page 28.)

In relativistic quantum field theory, the right-handed and left-handed electron and the right-handed and left-handed antielectron can be defined in terms of two fields denoted by e and e^c, where each field is a Weyl two-component left-handed spinor. The compositions of the fields are such that

e annihilates a left-handed electron e_L or creates a right-handed positron $\bar{e}_R$, and

e^c annihilates a left-handed positron $\bar{e}_L$ or creates a right-handed electron e_R.

These fields are complex, and for the action of the Hermitian conjugate fields $e^\dagger$ and $e^{c\dagger}$, just interchange the words annihilate and create above. For example, $e^\dagger$ creates a left-handed electron or annihilates a right-handed positron. Hence, the fields e and e^c and their complex conjugates can create or annihilate all the possible excitations of the physical electron. Note that parity (defined as the inversion

of spatial coordinates) has the property of interchanging the two states e_R and e_L.

What about the neutrino? The right-handed neutrino has never been observed, and it is not known whether that particle state and the left-handed antineutrino exist. In the Standard Model, the field $\nu_e{}^c$, which would create those states, is not included. Instead, the neutrino is associated with only two types of ripples (particle states) and is defined by a single field ν_e:

ν_e annihilates a left-handed electron neutrino ν_{eL} or creates a right-handed electron antineutrino $\bar{\nu}_{eR}$.

The left-handed electron neutrino has fermion number $N = +1$, and the right-handed electron antineutrino has fermion number $N = -1$. This description of the neutrino is not invariant under the parity operation. Parity interchanges left-handed and right-handed particles, but we just said that, in the Standard Model, the right-handed neutrino does not exist. The left-handedness of the neutrino mimics the left-handedness of the charged-current weak interactions. In other words, the W gauge boson, which mediates all weak charge-changing processes, acts only on the fields e and ν_e. The interaction with the W transforms the left-handed neutrino into the left-handed electron and vice versa ($e_L \leftrightarrow \nu_{eL}$) or the right-handed antineutrino into the right-handed positron and vice versa ($\bar{\nu}_{eR} \leftrightarrow \bar{e}_R$). Thus, we say that the fields e and ν_e, or the particles e_L and ν_{eL}, are a weak isospin doublet under the weak interactions.

These lepton fields carry two types of weak charge: The weak isotopic charge $I_3{}^w$ couples them to the W and the Z^0, and the weak hypercharge Y^w couples them to the Z^0. (The Z^0 is the neutral gauge boson that mediates neutral-current weak interactions.) Electric charge Q is related to the two weak charges through the equation $Q = I_3{}^w + Y^w/2$. Table I lists the weak charges for the particle states defined by the three fields e, ν_e, and e^c. Note that the particle states e_R and $\bar{e}_L$ defined by the field e^c do not couple to the W and have no weak isotopic charge. The field and the particle states are thus called weak isotopic singlets. However, e_R and $\bar{e}_L$ do carry weak hypercharge and electric charge and therefore couple to the Z^0 and the photon.

Likewise, the field $\nu_e{}^c$ and its neutrino states ν_R and $\bar{\nu}_L$ would be isotopic singlets with no coupling to the W. But unlike their electron counterparts, they must be electrically neutral ($Q = I_3{}^w + Y^w/2 = 0$), which implies they cannot have weak hypercharge. Thus, they would not couple to the W, the Z^0, or the photon. Having no interactions and, therefore, not being measurable, they are called *sterile* neutrinos and are not included in the Standard Model. However, if the left-handed neutrino has mass, it may oscillate into a sterile right-handed neutrino, a possibility that could be invoked in trying to give consistency to all the data on neutrino oscillations.

Table I. Lepton Charges

$$Q = I_3{}^w + \frac{Y^w}{2}$$

N	N_x	Particle States	$I_3{}^w$	Y^w	Q
$+1$	L	$\begin{pmatrix} e_L \\ \nu_L \end{pmatrix}$	$-1/2$	-1	-1
$+1$	L		$+1/2$	-1	0
-1	R	$\begin{pmatrix} \bar{e}_R \\ \bar{\nu}_R \end{pmatrix}$	$+1/2$	$+1$	$+1$
-1	R		$-1/2$	$+1$	0
$+1$	R	e_R	0	-2	-1
-1	L	$\bar{e}_L$	0	$+2$	$+1$
$+1$	R	ν_R	0	0	0
-1	L	$\bar{\nu}_L$	0	0	0

The Origin of Electron Mass in the Standard Model.
What is mass? Mass is the inertial energy of a particle. It is the energy a particle has when at rest and the measure of the resistance to an applied force according to Newton's law $F = ma$. A massless particle cannot exist at rest; it must always move at the speed of light.

A fermion (spin-1/2 particle) with mass has an additional constraint. It must exist in both right-handed and left-handed states because the only field operators that yield a nonzero mass for fermions are bilinear products of fields that flip the particle's handedness. For example, in the two-component notation introduced above, the standard, or Dirac, mass term in the Lagrangian for free electrons is given by[*]

$$m_e e^c e \ . \tag{1}$$

This fermion mass operator annihilates a left-handed electron and creates a right-handed electron in its place. The mass term does not change the charge of the particle, so we say that it conserves electric charge. Also, because this mass term does not change a particle into an antiparticle, we say that it conserves fermion number N. However, the weak isospin symmetry forbids such a mass operator because it is not an invariant under that symmetry. (The field e is a member of a weak isotopic doublet, whereas the field e^c is a weak isotopic singlet, so that the product of the two is not a singlet as it should be to preserve the weak isospin symmetry.) But the electron does have mass. We seem to be in a bind.

The Standard Model solves this problem: the electron and electron neutrino fields are postulated to interact with the spin-zero Higgs field h^0 (the God particle). The field h^0 is one member of a weak isospin doublet whose second member is h^+. The superscripts denote the electric charge of the state annihilated by each field (see Table III on page 57 for the other quantum number of the two fields). The field h^0 plays a special role in the Standard Model. Its ground state is not a vacuum state empty of particles, but it has a nonzero mean value, much like a Bose-Einstein condensate. This nonzero value, written as the vacuum expectation value $<0|h^0|0> \equiv <h^0> = v/\sqrt{2}$ is the putative "origin of mass." (The "mystery" of mass then becomes the origin of the Higgs boson and its nonzero vacuum value.)

The interaction between the Higgs fields and the electron and electron neutrino is given by

$$\lambda_e e^c \left(\nu_e (h^+)^\dagger + e(h^0)^\dagger \right) \ , \tag{2}$$

where λ_e is called a Yukawa coupling constant and describes the strength of the coupling between the Higgs field and the electron. The Higgs field is a weak isospin doublet, so the term in parentheses is an inner product of two doublets, making an invariant quantity under the weak isospin symmetry. Since it also conserves weak hypercharge, it preserves the symmetries of the Standard Model.

Because the mean value of h^0 in the vacuum is $<h^0> = v/\sqrt{2}$, the operator in (2) contributes a term to the Standard Model of the form

$$\lambda_e <h^0> e^c e = (\lambda_e v/\sqrt{2}) e^c e \ . \tag{3}$$

In other words, as the electron moves through the vacuum, it constantly feels the interaction with the Higgs field in the vacuum. But (3) is a fermion mass operator exactly analogous to the Dirac mass operator in (1), except that here the electron rest mass is given by

$$m_e = \lambda_e v/\sqrt{2} \ . \tag{4}$$

We see that, in the Standard Model, electron mass comes from the Yukawa interaction of the electron with the Higgs background.

Why Neutrinos Are Massless in the Minimal Standard Model. What about the neutrino? Because the neutrino has spin 1/2, its mass operator must also change handedness if it is to yield a nonzero value. We could introduce a Dirac mass term

---

[*] The addition of the Hermitian conjugate is assumed in all equations if the operator is not explicitly Hermitian.

for the neutrino that would mirror the mass term for the electron. It would
have the form

$$m_\nu \nu_e{}^c \nu_e \ .$$ (5)

But, as we said above, the field $\nu_e{}^c$ is not included in the Standard Model
because, so far, weak-interaction experiments have not required it. The neutrino,
though, has no electric charge, which makes it possible to write down a mass term
from the existing neutrino field ν_e with the form

$$\frac{1}{2} \, \mu_\nu \nu_e \nu_e \ .$$ (6)

(Note that m_ν and μ_ν refer just to the electron neutrinos, but similar masses can
be defined for the μ and τ neutrinos.) The mass operator in (6) annihilates a left-
handed neutrino and creates a right-handed antineutrino, which means that it is a
Majorana mass term. *Any mass term that changes a particle to an antiparticle is
called a Majorana mass term.* In changing a neutrino to an antineutrino, this term
violates fermion number N, changing it by two units. It is a legitimate mass term
in that it changes handedness in the right way to yield a nonzero rest mass, and it
conserves electric charge because the neutrino is electrically neutral. Nevertheless,
it is not included in the Standard Model because it violates the weak symmetry in
two ways: It is not invariant under the weak isospin symmetry, and it changes the
weak hypercharge by two units. We conclude that, in the minimal Standard Model,
which does not include $\nu_e{}^c$ and contains only the Higgs doublet mentioned above,
there is no way to give mass to the neutrinos if fermion number is conserved.

Two consequences follow directly from the result that neutrino masses are
identically zero in the minimal Standard Model. First, the weak eigenstates and
the mass eigenstates of the leptons are equivalent, and therefore individual-lepton-
family number (electron number, muon number, and tau number) are conserved
(for the proof, see "Family Mixing and the Origin of Mass" on page 72). Thus, the
Standard Model forbids such processes as

$$\mu^+ \to e^+ + \gamma \ , \text{ or}$$ (7)

$$\mu^+ \to e^+ + e^+ + e^- \ .$$ (8)

Similarly, the proposed process of neutrino oscillation, which may recently have
been observed, is forbidden. Second, total lepton number, equal to the sum
of individual-family-lepton numbers, is also conserved, and the process of
neutrinoless double beta decay is forbidden.

The converse is also true: If individual-lepton-number violation is observed,
or if the LSND results on neutrino oscillation are confirmed, then either of those
experiments could claim the discovery of nonzero neutrino masses and thus of
new physics beyond the Standard Model.

Adding Neutrino Masses to the Standard Model. What could this new
physics be? There are several simple extensions to the Standard Model that
could yield nonzero neutrino masses without changing the local symmetry of
the weak interactions.

The simplest extension would be to add no new fields but just a new
"effective" interaction with the Higgs field:

$$\frac{1}{M_{\text{effective}}}(h^0 \nu_e - h^+ e)^2 \ .$$ (9)

This effective interaction is invariant under the local symmetries and yields a Majorana mass term equal to

$$\frac{1}{M_{\text{effective}}} <h^0>^2 \nu_e \nu_e \; , \tag{10}$$

and a value for the neutrino mass

$$\mu_\nu = \frac{2<h^0>^2}{M_{\text{effective}}} = \frac{\text{v}^2}{M_{\text{effective}}} \; . \tag{11}$$

This mass term, as all fermion mass terms, changes handedness from left to right, but it violates the fermion number N listed in Table I. The term $M_{\text{effective}}$ must be large so that the mass of the neutrino be small. The new term in (9) is called "effective" because it can only be used to compute the physics at energies well below $M_{\text{effective}}c^2$, just as Fermi's "effective" theory of beta decay yields valid approximations to weak processes only at energies well below $M_W c^2$, where M_W is the mass of the W. (Outside their specified energy ranges, "effective" theories are, in technical language, nonrenormalizable and yield infinite values for finite quantities.) Thus, the mass term in (9) implicitly introduces a new scale of physics, in which new particles with masses on the order of $M_{\text{effective}}$ presumably play a role. Below that energy scale, (9) describes the effects of the seesaw mechanism for generating small neutrino masses (see below as well as the box "The Seesaw Mechanism at Low Energies" on page 71).

A Dirac Mass Term. Another extension would be to introduce a right-handed neutrino field ν_i^c, one for each neutrino flavor i ($i = e, \mu, \tau$), where, for example, the right-handed field for the electron neutrino is defined such that

ν_e^c annihilates a left-handed electron antineutrino $\overline{\nu}_{eL}$ and creates a right-handed electron neutrino ν_{eR}.

We could then define an interaction with the Higgs field exactly analogous to the interaction in (3) that gives electrons their mass:

$$\lambda_\nu \, \nu_e^c (\nu h^0 - e h^+) \; . \tag{12}$$

Again, because the Higgs field h^0 has a nonzero vacuum expectation value, the interaction in (12) would give the neutrino a Dirac mass

$$m_\nu = \frac{\lambda_\nu \text{v}}{\sqrt{2}} \; . \tag{13}$$

But why are neutrino masses much smaller than the masses of their charged lepton weak partners? Specifically, why is $m_\nu << m_e$? The electron mass is 500,000 eV, whereas from experiment, the electron neutrino mass is known to be less than 10 eV. The only explanation within the context of the interaction above is that the strength of the Yukawa coupling to the Higgs field is much greater for the electron than for the electron neutrino, that is, $\lambda_e > 5 \times 10^4 \lambda_\nu$. But this is not an explanation; it just parametrizes the obvious.

The Seesaw Mechanism and Majorana Neutrinos. The first real model of why neutrino masses are very much smaller than the masses of their lepton partners was provided by Murray Gell-Mann, Pierre Ramond, and Richard Slansky. Motivated by a class of theories that attempt to unify the interactions of the

Standard Model, including the strong interactions, they observed that, if one introduced the right-handed neutrino field $\nu_e{}^c$ into the Standard Model to form a Dirac mass term, one could also add a Majorana mass term of the form

$$\frac{1}{2}M\nu_e{}^c\nu_e{}^c \tag{14}$$

without violating the local symmetries of the Standard Model (as stated above, $\nu_e{}^c$ has no weak charge and is thus an invariant under the local symmetry). Further, if M were large enough, the mass of the left-handed neutrino would be small enough to satisfy the experimental bounds.

To see how this reduction occurs, we write the operators for both the Dirac mass term and the Majorana mass term:

$$\mathscr{L}_{\text{mass}} = \lambda_\nu (h^0\nu_e - h^+e)\nu_e{}^c + \frac{1}{2}M\nu_e{}^c\nu_e{}^c + \text{other terms} \ . \tag{15}$$

Here we are assuming that $\lambda_\nu \cong \lambda_e$. These additions to the Lagrangian yield the following mass terms:

$$\mathscr{L}_{\nu_e \text{ mass}} = m_{\nu_e}\nu_e\nu_e{}^c + \frac{1}{2}M\nu_e{}^c\nu_e{}^c \ , \tag{16}$$

where m_{ν_e} is the Dirac mass defined in (13), except that now we assume $\lambda_\nu \cong \lambda_e$, in which case $m_{\nu_e} \cong \lambda_e v/\sqrt{2}$. In other words, the Dirac neutrino mass is about equal to the electron mass (or some other fermion mass in the first family). The two neutrino mass terms may be rewritten as a matrix, frequently referred to as the mass matrix:

$$1/2(\nu_e \quad \nu_e{}^c)\begin{pmatrix} 0 & m_{\nu_e} \\ m_{\nu_e} & M \end{pmatrix}\begin{pmatrix} \nu_e \\ \nu_e{}^c \end{pmatrix} \ . \tag{17}$$

It is clear that the fields ν_e and $\nu_e{}^c$ do not describe states of definite mass, or mass eigenstates, but rather the two fields are mixed by the interaction with the Higgs field. Diagonalizing this matrix yields the masses of the physical neutrinos. [The expressions in (16) and Equation (17) are equivalent. The proof requires more detail than is presented here.] One mass is very small:

$$\mu_{\text{light}} \approx \frac{m_{\nu_e}^2}{M} \ . \tag{18}$$

It is the Dirac mass reduced by ratio m_{ν_e}/M that gave this mechanism its name—the "seesaw." The second mass is very large:

$$\mu_{\text{heavy}} \approx M \ . \tag{19}$$

The fields corresponding to these masses are given by

$$\nu_{\text{light}} \approx \nu_e + \left(\frac{m_{\nu_e}}{M}\right)\nu_e{}^c \approx \nu_e \ , \tag{20}$$

and

$$\nu_{\text{heavy}} \approx \nu_e{}^c - \left(\frac{m_{\nu_e}}{M}\right)\nu_e \approx \nu_e{}^c \ . \tag{21}$$

Both fields define Majorana particles, that is, particles that are their own antiparticles, and total-lepton-number conservation can be violated in processes involving these neutrinos. The light neutrino would correspond to the neutrino we see in the

weak processes observed so far, and is essentially the left-handed neutrino field ν_e. The right-handed neutrino field $\nu_e{}^c$ would not be observed directly at low energies. Its effect in the low-energy theory would only be visible as an effective neutrino mass operator, like the operator in (9), which would give the neutrino a very small mass and would signal the presence of a new scale of physics on the order of $M_{\text{effective}} = 2M/\lambda_\nu{}^2$ (see the box "The Seesaw Mechanism at Low Energies" on the facing page).

A New Higgs Isospin Triplet. Another possibility is that there are no right-handed neutrinos, but there is, instead, a new set of Higgs-type bosons ϕ that come in three varieties —ϕ^0, ϕ^+, ϕ^{++}— and transform as a triplet under the local weak isospin symmetry. The superscript denotes the electric charge of each boson. Using this Higgs triplet, we can introduce the interaction

$$\lambda_m(\nu\nu\phi^0 + \nu e\phi^+ + ee\phi^{++}) \ , \qquad (22)$$

which is consistent with all Standard Model symmetries. If, in analogy with h^0, the Higgs field ϕ^0 has a nonzero vacuum expectation value $<\phi^0> = \text{v}_m$, the neutrino would also have a Majorana mass given by

$$\mu_\nu = \lambda_m<\phi^0> = \lambda_m\text{v}_m \ , \qquad (23)$$

where this fermion mass is a Majorana mass. In a theory with a Higgs triplet, the Higgs doublet is still necessary. In fact, in order to preserve the observed ratio of strengths of neutral- to charged-current interactions (equal to $1 \pm .01$), the vacuum expectation value v_m must be much smaller than in (3). Also, such a theory has a massless Nambu-Goldstone boson ϕ due to the spontaneous breaking of total lepton number, and it allows the process

$$\nu_\mu \rightarrow \nu_e + \phi \ . \qquad (24)$$

Apart from the effective interaction in Equation (9), the other extensions we discussed introduce new states. Each makes predictions that can be tested. The Higgs triplet extension is the largest departure from the Standard Model. The seesaw mechanism is less intrusive than the Higgs triplet. In general, its only low-energy consequence is an arbitrary Majorana mass term for the three neutrino species given by

$$\mu_{ij}\nu_i\nu_j \ , \text{ where } i, j = e, \mu, \tau \ . \qquad (25)$$

A general mass matrix such as the one in (25) would lead to lepton-family-number violating processes, CP (charge-conjugation/parity) violation, and neutrino oscillations. This simple hypothesis will be tested by present or proposed experiments.

On a final note, the new scale M in (15) can be very large. It may be associated with the proposed grand unification scale for strong, weak, and electromagnetic interactions, which is predicted to occur at energies on the order of 10^{16} GeV. If so, neutrino masses and mixings can give us information about the physics at this enormous energy scale. There is also the exciting possibility that, through a sequence of interactions that violate CP, lepton-number, and baryon-number conservation, the decay of the very heavy right-handed neutrino ν^c in the hot, early universe generates the observed baryon number of the universe, that is, the presence of matter as opposed to antimatter. ∎

The Seesaw Mechanism at Low Energies

The seesaw mechanism for neutrino masses defines a new scale of nature given by M, the mass associated with the heavy right-handed neutrino ν_e^c. Since M is postulated to be very large, well above the energies accessible through experiment, it is interesting that the "effective" neutrino mass operator in (11) approximates the seesaw terms in (15) at energies below M. To show this, we consider the effective operator

$$\frac{1}{M_{\text{effective}}}(h^0\nu_e - h^+e)^2 \ .$$

When the Higgs vacuum expectation value is accounted for, this operator yields the nonrenormalizable mass term in diagram (a) and a Majorana mass given by

$$\mu_\nu = \frac{v^2}{M_{\text{effective}}}$$

In the seesaw mechanism, the light neutrino acquires its mass through the exchange of the heavy neutrino, as shown in diagram (b). Diagram (b), which is approximated by diagram (a) at energies below Mc^2, is a renormalizable mass term that involves both Dirac and Majorana masses. It yields a neutrino mass

$$\mu_{\text{light}} = \frac{m_{\nu_e}^2}{M} \quad \text{with} \quad m_{\nu_e} \equiv \lambda_{\nu_e}\frac{v}{\sqrt{2}} \ .$$

Equating the values for μ_ν and μ_{light}, we obtain the relation between M and $M_{\text{effective}}$:

$$\frac{1}{M_{\text{effective}}} = \frac{(\lambda_{\nu_e})^2}{2M} \ .$$

At energies below M_W, the mass of the W boson, a similar type of relationship exists between Fermi's "effective" theory shown in diagram (c) and the W-boson exchange processes shown in diagram (d). The exchange processes are defined by the gauge theory of the charged-current weak interactions. Fermi's theory is a nonrenormalizable current-current interaction of the form

$$\mathscr{L}_{\text{Fermi}} = \frac{G_F}{\sqrt{2}}J^\mu_W{}^\dagger J_\mu^W \ ,$$

where the weak current for the neutrino-electron doublet is given by

$$J^\mu_W = 2\nu_e^\dagger \overline{\sigma}^\mu e \quad \text{and} \quad \overline{\sigma}^\mu = (1, -\sigma^i) \ ,$$

and the Fermi constant G_F defines the strength of the effective interaction in diagram (c), as well as a new mass/energy scale of nature. The experimentally observed value is $G_F = 1.66 \times 10^{-5} \text{ GeV}^{-2}$. Equating the low-energy limit of diagram (c) with that of diagram (d) yields the formula

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2} \ ,$$

where g is the weak isospin coupling constant in the charged-current weak Lagrangian given by

$$\mathscr{L}_{\text{weak}} = -M_W^2 W^{\mu+}W_\mu^- + \frac{g}{2\sqrt{2}}W_\mu^+ J^\mu_W + \frac{g}{2\sqrt{2}}W_\mu^- J^{\mu\dagger}_W \ .$$

This Lagrangian neglects the kinetic term for the W, which is a valid approximation at energies much less than the W boson mass.

(a) Effective neutrino mass term
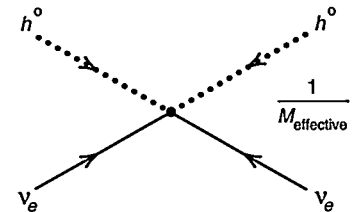
(b) Seesaw mass term for the light neutrino
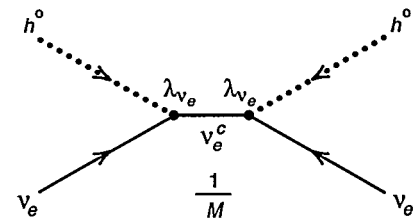
(c) Fermi's current-current interaction
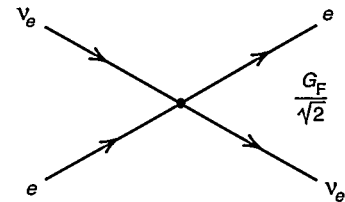
(d) Weak charged-current gauge interaction
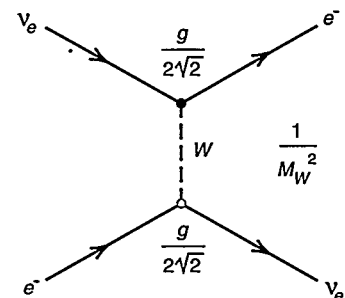
Family Mixing and the Origin of Mass

The difference between weak eigenstates and mass eigenstates

Stuart Raby

The Standard Model of elementary particle physics contains two disjoint sectors. The gauge sector describes the interactions of quarks and leptons (fermions, or spin-1/2 particles) with the spin-1 gauge bosons that mediate the strong, weak, and electromagnetic forces. This sector has great aesthetic appeal because the interactions are derived from local gauge symmetries. Also, the three families of quarks and leptons transform identically under those local symmetries and thus have the same basic strong, weak, and electromagnetic interactions.

The Higgs sector describes the interactions of the quarks and leptons with the spin-0 Higgs bosons h^+ and h^0. This sector is somewhat ad hoc and contains many free parameters. The Higgs bosons were originally introduced to break the weak isospin gauge symmetry of the weak interactions by giving mass to the weak gauge bosons, the W and the Z^0. The W and the Z^0 must be very heavy to explain why the weak force is so weak. But in the Standard Model, interactions with those Higgs bosons are also responsible for giving nonzero masses to the three families of quarks and leptons. Those interactions must yield different masses for the particles from different families and must cause the quarks from different families to mix, as observed in experiment. But neither the nine masses for the quarks and charged leptons nor the four parameters that specify the mixing of quarks across families are determined by any fundamental principle contained in the Standard Model. Instead, those thirteen parameters are determined from low-energy experiments and are matched to the free parameters in the Standard Model Lagrangian.

By definition, weak eigenstates are the members of the weak isospin doublets that transform into each other through interaction with the W boson (see Figure 5 on page 38). Mass eigenstates are states of definite mass created by the interaction with Higgs bosons. Those states describe freely propagating particles that are identified in detectors by their electric charge, mass, and spin quantum numbers. Since the Higgs interactions cause the quark weak eigenstates to mix with each other, the resulting mass eigenstates are not identical to the weak eigenstates.

Each set of eigenstates provides a description of the three families of quarks, and the two descriptions are related to each other by a set of unitary rotations. Most experimentalists are accustomed to seeing the Standard Model written in the mass eigenstate basis because the quarks of definite mass are the ingredients of protons, neutrons, and other metastable particles that the experimentalists measure. In the mass eigenstate basis, the Higgs interactions are diagonal, and the mixing across families appears in the gauge sector. In other words, the unitary rotations connecting the mass eigenstate basis to the weak eigenstate basis appear in the gauge interactions. Those rotation matrices could, in principle, appear in all the gauge interactions of quarks and leptons; but they do not. The Standard Model symmetries cause the rotation matrices to appear only in the quark charge-changing currents that couple to the W boson.

The specific product of rotation matrices that appears in the weak charge-changing currents is just what we call the CKM matrix, the unitary 3×3 mixing matrix deduced by Cabibbo, Kobayashi, and Maskawa. The elements in the CKM matrix have been determined by measuring, for example, the strengths of the strangeness-changing processes, in which a strange quark from the second family of mass states transforms into an up quark from the first family. So far, family mixing has not been observed among the leptons, with the possible

exception of neutrino oscillations. If oscillations are confirmed, the mixing angles measured in the neutrino experiments will become part of a CKM mixing matrix for the leptons.

This sidebar derives the form of the CKM matrix and shows how it reflects the difference between the rotation matrices for the up-type quarks ($Q = +2/3$) and those for their weak partners, the down-type quarks ($Q = -1/3$). This difference causes the family mixing in weak-interaction processes and is an example of the way in which the Higgs sector breaks the weak symmetry. We will also show that, because the neutrino masses are assumed to be degenerate (namely, zero), in the Standard Model, the rotation matrices for the neutrinos can be defined as identical to those for their weak partners, and therefore the CKM matrix for the leptons is the identity matrix. Thus, in the minimal Standard Model, in which neutrinos are massless, no family mixing can occur among the leptons, and individual-lepton-family number is conserved.

This discussion attributes the origin of mixing to the mismatch between weak eigenstates and mass eigenstates caused by the Higgs sector. A more fundamental understanding of mixing would require understanding the origin of fermion masses and the reason for certain symmetries, or approximate symmetries, to hold in nature. For example, a fundamental theory of fermion masses would have to explain why muon-family number is conserved, or only approximately conserved. It would also have to explain why the $K^0 - \overline{K}^0$ mixing amplitude is on the order of G_F^2 and not larger. The small amount of family mixing observed in nature puts severe constraints on any theory of fermion masses. Developing such a theory is an outstanding problem in particle physics, but it may require a significant extension of the Standard Model.

To discuss mixing as it appears in the Standard Model, it is necessary to explicitly write down the parts of the Standard Model Lagrangian that contain the Yukawa interactions between the fermions and the Higgs bosons (responsible for fermion masses) and the weak gauge interaction between the fermions and the W boson (responsible for charge-changing processes such as beta decay). But first, we must define some notation. As in the sidebar "Neutrino Masses" on page 64, we describe the fermion states by two-component left-handed Weyl spinors. Specifically, we have the fields u_i, d_i, u_i^c, d_i^c, e_i, v_i, and e_i^c, where the family index i runs from one to three. The u_i are the fields for the three up-type quarks u, c, and t with electric charge $Q = +2/3$, the d_i are the fields for the three down-type quarks d, s, and b with $Q = -1/3$, the e_i stand for the three charged leptons e, μ, and τ with $Q = -1$, and the v_i stand for the three neutrinos v_e, v_μ, and v_τ with $Q = 0$. The fields u_i and u_i^c, for example, are defined as follows:

u_i annihilates the left-handed up-type quark u_L and creates the right-handed up-type antiquark $\overline{u}_R$ in family i, and

u_i^c annihilates the left-handed up-type antiquark $\overline{u}_L$ and creates the right-handed up-type quark u_R in family i.

To describe the Hermitian conjugate fields $u_i^\dagger$ and $u_i^{c\dagger}$, interchange the words annihilate and create used above. Thus u_i, u_i^c, and their Hermitian conjugates describe the creation and annihilation of all the states of the up-type quarks. The down-type quark fields and the charged lepton fields are similarly defined. For the neutrinos, only the fields v_i containing the states v_L and $\overline{v}_R$ are observed; the fields v_i^c are not included in the Standard Model. In other words, the Standard Model includes right-handed charged leptons, but it has no right-handed neutrinos (or left-handed antineutrinos).

The Pauli Matrices for Spin-1/2 Particles

The Pauli spin matrices generate all rotations of spin-1/2 particles. Spin-1/2 particles have only two possible spin projections along, say, the 3-axis: spin up, or $s_3 = +1/2$, and spin down, or $s_3 = -1/2$. The step-up operator σ^+ raises spin down to spin up, the step-down operator σ^- lowers spin up to spin down, and σ_3 gives the value of the spin projection along the 3-axis. The basis set for the spin quantized along the 3-axis is given by

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 1 \end{pmatrix} ,$$

and the matrices are given by

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

$$\sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} .$$

Defining the matrices $\sigma^\pm$ as

$$\sigma^\pm = \frac{1}{2}(\sigma^1 \pm i\sigma^2) ,$$

one arrives at the following commutation relations:

$$[\sigma^3, \sigma^\pm] = \pm 2\sigma^\pm , \text{ and}$$
$$[\sigma^+, \sigma^-] = \sigma^3 .$$

The Weak Eigenstate Basis. We begin by defining the theory in terms of the weak eigenstates denoted by the subscript 0 and the color red. Specifically, the weak gauge coupling to the W is given by

$$\mathscr{L}_{\text{weak}} = +\frac{g}{\sqrt{2}}(W_\mu^+ J^\mu + W_\mu^- J^{\mu\dagger}) , \qquad (1)$$

where the charge-raising weak current J^μ is defined as

$$J^\mu = \sum_i u_{0i}^\dagger \bar{\sigma}^\mu d_{0i} + \nu_{0i}^\dagger \bar{\sigma}^\mu e_{0i} , \qquad (2)$$

and the charge-lowering current $J^{\mu\dagger}$ is defined as

$$J^{\mu\dagger} = \sum_i d_{0i}^\dagger \bar{\sigma}^\mu u_{0i} + e_{0i}^\dagger \bar{\sigma}^\mu \nu_{0i} . \qquad (3)$$

The constant g in Equation (1) specifies the strength of the weak interactions, and the $\bar{\sigma}^\mu$ is a four-component space-time vector given by $(1, -\sigma^j)$, where the σ^j are the standard Pauli spin matrices for spin-1/2 particles with $j = x, y, z$, the spatial directions. These 2×2 matrices act on the spin components of the spin-1/2 fields and are totally independent of the family index i. Each term in the charge-raising and charge-lowering currents connects states from the same family, which means the weak interactions in Equation (1) are diagonal in the weak eigenstate basis. In fact, those interactions define the weak eigenstates.

To understand the action of the currents, consider the first term, $u_0^\dagger \bar{\sigma}^\mu d_0$, in the charge-raising current J^μ. It annihilates a left-handed down quark and creates a left-handed up quark ($d_{0L} \to u_{0L}$) and, thereby, raises the electric charge by one unit. Electric charge is conserved because the W^+ field creates a W^- (see top diagram at right). The first term in the charge-lowering current $J^{\mu\dagger}$ does the reverse: $d_0^\dagger \bar{\sigma}^\mu u_0$ annihilates a left-handed up quark and creates a left-handed down quark ($u_{0L} \to d_{0L}$) and, thereby, lowers the electric charge by one unit; at the same time, the W^- field creates a W^+ (see bottom diagram at right). Thus, the members of each pair u_{0i} and d_{0i} transform into each other under the action of the charge-raising and charge-lowering weak currents and therefore are, by definition, a weak isospin doublet. The quark doublets are (u_0, d_0), (c_0, s_0), and (t_0, b_0), and the lepton doublets are (ν_{e0}, e_0), $(\nu_{\mu 0}, \mu_0)$, and $(\nu_{\tau 0}, \tau_0)$. The first member of the doublet has weak isotopic charge $I_3^W = +1/2$, and the second member has $I_3^W = -1/2$.

Finally, note that J^μ and $J^{\mu\dagger}$ are left-handed currents. They contain only the fermion fields f_0 and not the fermion fields f_0^c, which means that they create and annihilate only left-handed fermions f_{0L} (and right-handed antifermions $\bar{f}_{0R}$). The right-handed fermions f_{0R} (and left-handed antifermions $\bar{f}_{0L}$) are simply impervious to the charge-changing weak interactions, and therefore, the f_0^c are weak isotopic singlets. They are invariant under the weak isospin transformations.

Weak isospin symmetry, like strong isospin symmetry from nuclear physics and the symmetry of rotations, is an SU(2) symmetry, which means that there are three generators of the group of weak isospin symmetry transformations. Those generators have the same commutation relations as the Pauli spin matrices. (The Pauli matrices, shown at left, generate all the rotations of spin-1/2 particles.) The J^μ and $J^{\mu\dagger}$ are the raising and lowering generators of weak isospin analogous to σ^+ and σ^-. The generator analogous to $1/2\sigma_3$ is J_3^μ given by

$$J_3^\mu = 1/2 \sum_i u_{0i}^\dagger \bar{\sigma}^\mu u_{0i} - d_{0i}^\dagger \bar{\sigma}^\mu d_{0i} + \nu_{0i}^\dagger \bar{\sigma}^\mu \nu_{0i} - e_{0i}^\dagger \bar{\sigma}^\mu e_{0i} , \qquad (4)$$

and the time components of these three currents obey the commutation relations $[J^0, J^{0\dagger}] = 2J_3^0$. In general, the time component of a current is the charge

density, whereas the spatial component is the flux. Similarly, $J_3{}^0$ is the weak isotopic charge density. It contains terms of the form $f_0{}^\dagger f_0$, which are number operators N_f that count the number of f particles minus the number of $\bar{f}$ antiparticles present. When this density is integrated over all space, it yields the weak isotopic charge $I_3{}^W$:

$$\int J_3{}^0(x)d^3x = I_3{}^W \ .$$

Now, let us consider the Higgs sector. The fermion fields interact with the Higgs weak isospin doublet (h^+, h^0) through the Yukawa interactions given by

$$\mathcal{L}_{\text{Yukawa}} = \sum_{i,j} u^c{}_{0i} \ (Y_{\text{up}} \ )_{ij} \ [u_{0j} \ h^0 - d_{0j} \ h^+] + d^c{}_{0i} \ (Y_{\text{down}})_{ij} \ [u_{0j} \ (h^+)^\dagger + d_{0j} \ (h^0)^\dagger] + e^c{}_{0i} \ (Y_{\text{lepton}})_{ij} \ [\nu_{0j} \ (h^+)^\dagger + e_{0j} \ (h^0)^\dagger] \ \ , \ (5)$$

where Y_{up}, Y_{down}, and Y_{lepton} are the complex 3×3 Yukawa matrices that give the strengths of the interactions between the fermions and the Higgs bosons. Because the Higgs fields form a weak isospin doublet, each expression in brackets is an inner product of two weak doublets, making an isospin singlet. Thus, each term in the Lagrangian is invariant under the local weak isospin symmetry since the conjugate fields (for example, $u^c{}_0$) are weak singlets. The lepton terms in Equation (5) are introduced in the sidebar "Neutrino Masses" (page 64), where masses are shown to arise directly from the Yukawa interactions because h^0 has a nonzero vacuum expectation value $<h^0> = v/\sqrt{2}$ that causes each type of fermion to feel an everpresent interaction. These interactions yield mass terms given by

$$\mathcal{L}_{\text{Yukawa}} \rightarrow \mathcal{L}_{\text{mass}} = u^c{}_{0i} \ (Y_{\text{up}} \ )_{ij} \ u_{0j} <h^0> + d^c{}_{0i} \ (Y_{\text{down}})_{ij} \ d_{0j} <h^{0\dagger}> + e^c{}_{0i} \ (Y_{\text{lepton}})_{ij} \ e_{0j} <h^{0\dagger}> \ . \ (6)$$

Notice that each term in $\mathcal{L}_{\text{mass}}$ contains a product of two fermion fields $f_0{}^c f_0$, which, by definition, annihilates a left-handed fermion and creates a right-handed fermion. Thus, these Yukawa interactions flip the handedness of fermions, a prerequisite for giving nonzero masses to the fermions. These terms resemble the Dirac mass terms introduced in the sidebar "Neutrino Masses," except that the matrices Y_{up}, Y_{down}, and Y_{lepton} are *not* diagonal. Thus, in the weak eigenstate basis, the masses and the mixing across families occur in the Higgs sector.

The Mass Eigenstate Basis and the Higgs Sector. Let us examine the theory in the mass eigenstate basis. We find this basis by diagonalizing the Yukawa matrices in the mass terms of Equation (6). In general, each Yukawa matrix is diagonalized by two unitary 3×3 transformation matrices. For example, the diagonal Yukawa matrix for the up quarks $\hat{Y}_{\text{up}}$ is given by

$$\hat{Y}_{\text{up}} = V_u^R Y_{\text{up}} V_u^{L\dagger} \ , \ (7)$$

where matrix V_u^R acts on the right-handed up-type quarks in the fields $u^c{}_0$, and matrix V_u^L acts on the left-handed up-type quarks in u_0. The diagonal elements of $\hat{Y}_{\text{up}}$ are $(\lambda_u, \lambda_c, \lambda_t \ )$, the Yukawa interaction strengths for all the up-type quarks: the up, charm, and top, respectively. Matrices $\hat{Y}_{\text{down}}$ and $\hat{Y}_{\text{lepton}}$ are similarly diagonalized. If u_0 and $u^c{}_0$ are the fields in the weak eigenstate, the fields in the mass eigenstate, u^c and u, are defined by the unitary transformations

$$u^c{}_0 = u^c \ V_u^R \ \text{and} \ u_0 = V_u^{L\dagger} \ u \ . \ (8)$$

Since the Vs are unitary transformations, $V^\dagger V = V V^\dagger = I$, we also have

The charge-raising weak interaction in the first family

$$(W_\mu^+ \ J^\mu)_{\text{first family}} = W_\mu^+ \ u_0{}^\dagger \ \bar{\sigma}^\mu \ d_0 \ .$$
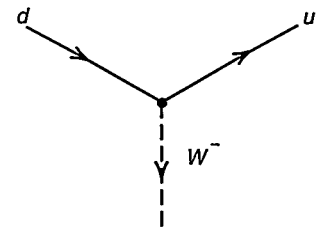
A down quark changes to an up quark with the emission of a W^-.

The charge-lowering weak interaction in the first family

$$(W_\mu^- \ J^{\mu\dagger})_{\text{first family}} = W_\mu^- \ d_0{}^\dagger \ \bar{\sigma}^\mu \ u_0 \ .$$
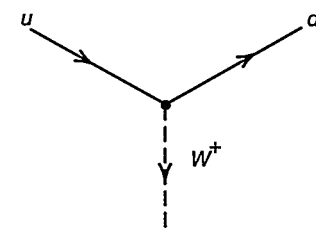
An up quark changes to a down quark with the emission of a W^+.

$$u^c = u^c_0 \, V_u^{R\,\dagger} \quad \text{and} \quad u = V_u^L u_0 \ .$$

In this new mass basis, $\mathscr{L}_{\text{mass}}$ in Equation (6) takes the form

$$\mathscr{L}_{\text{mass}} = \sum_i u^c{}_i \, \hat{Y}^i_{\text{up}} \, u_i <h^0> + d^c_i \, \hat{Y}^i_{\text{down}} \, d_i <h^0> + e^c{}_i \, \hat{Y}^i_{\text{lepton}} \, e_i <h^0>$$

$$= \sum_i u^c{}_i \, \hat{M}^i_{\text{up}} \, u_i + d^c_i \, \hat{M}^i_{\text{down}} \, d_i + e^c{}_i \, \hat{M}^i_{\text{lepton}} \, e_i \ , \tag{9}$$

where the matrices $\hat{M}^i = \hat{Y}^i \text{v}/\sqrt{2}$ are diagonal, and the diagonal elements are just the masses of the fermions. In particular, we can write out the three terms for the up-type quarks u, c, and t:

$$\sum_i u^c{}_i \, \hat{M}^i_{\text{up}} \, u_i = \lambda_u \, u^c \, u <h^0> + \lambda_c \, c^c \, c <h^0> + \lambda_t \, t^c \, t <h^0>$$

$$= \lambda_u \, \text{v}/\sqrt{2} u^c \, u + \lambda_c \, \text{v}/\sqrt{2} \, c^c \, c + \lambda_t \, \text{v}/\sqrt{2} \, t^c \, t$$

$$= m_u \, u^c \, u + m_c \, c^c \, c + m_t \, t^c \, t \ , \tag{10}$$

with the masses of the up, charm, and top quarks given by

$$m_u = \lambda_u \text{v}/\sqrt{2}, \quad m_c = \lambda_c \text{v}/\sqrt{2}, \quad \text{and } m_t = \lambda_t \text{v}/\sqrt{2} \ .$$

Thus, the Higgs sector defines the mass eigenstate basis, and the diagonal elements of the mass matrices are the particle masses.

Mixing in the Mass Eigenstate Basis. Now, let us write the weak gauge interaction with the W in the mass eigenstate. Recall that

$$\mathscr{L}_{\text{weak}} = + \frac{g}{\sqrt{2}} (W^+_\mu J^\mu + W^-_\mu J^{\mu\dagger}) \ ,$$

but to write the charge-raising weak current J^μ in the mass eigenstate, we substitute Equation (8) into Equation (2),

$$J^\mu = \sum_i u_{0i}{}^\dagger \overline{\sigma}^\mu d_{0i} + \nu_{0i}{}^\dagger \overline{\sigma}^\mu e_{0i}$$

$$= \sum_{i,k,j} u_i{}^\dagger (V_u^L)_{ik} \overline{\sigma}^\mu (V_d^{L\dagger})_{kj} d_j + \nu_i{}^\dagger \overline{\sigma}^\mu e_i$$

$$= \sum_{i,j} u_i{}^\dagger \overline{\sigma}^\mu (V_{\text{CKM}})_{ij} d_j + \nu_i{}^\dagger \overline{\sigma}^\mu e_i \ , \tag{11}$$

and to rewrite the charge-lowering current $J^{\mu\dagger}$, we substitute Equation (8) into Equation (3):

$$J^{\mu\dagger} = \sum_i d_{0i}{}^\dagger \overline{\sigma}^\mu u_{0i} + e_{0i}{}^\dagger \overline{\sigma}^\mu \nu_{0i}$$

$$= \sum_{i,k,j} d_i{}^\dagger (V_d^L)_{ik} \overline{\sigma}^\mu (V_u^{L\dagger})_{kj} u_j + e_i{}^\dagger \overline{\sigma}^\mu \nu_i$$

$$= \sum_{i,j} d_i{}^\dagger \overline{\sigma}^\mu (V_{\text{CKM}})^\dagger{}_{ij} u_j + e_i{}^\dagger \overline{\sigma}^\mu e_i \ , \tag{12}$$

where

$$V_{\text{CKM}} = V_u^L \, V_d^{L\dagger} \ . \tag{13}$$

Thus, the charge-raising and charge-lowering quark currents are *not* diagonal in the mass eigenstate basis. Instead, they contain the complex 3×3 mixing matrix V_{CKM}. This matrix would be the identity matrix were it not for the

difference between the rotation matrices for the up-type quarks V_u^L and those for the down-type quarks V_d^L. It is that difference that determines the amount of family mixing in weak-interaction processes. For that reason, all the mixing can be placed in either the up-type or down-type quarks, and by convention, the CKM matrix places all the mixing in the down-type quarks. The weak eigenstates for the down-type quarks are often defined as d':

$$d' = V_{\text{CKM}} d = V_u^L V_d^{L\dagger} d = V_u^L d_0 \ , \qquad (14)$$

in which case, the up-type weak partners to d' become u':

$$u' = V_u^L u_0 \equiv u \ .$$

When all the mixing is placed in the down-type quarks, the weak eigenstates for the up-type quarks are the same as the mass eigenstates. (We could just as easily place the mixing in the up-type quarks by defining a set of fields u' given in terms of the mass eigenstates u and V_{CKM}.) Independent of any convention, the weak currents J^μ couple quark mass eigenstates from different families. The form of the CKM matrix shows that, from the Higgs perspective, the up-type and down-type quarks look different. It is this mismatch that causes the mixing across quark families. If the rotation matrices for the up-type and down-type left-handed quarks were the same, that is, if $V_u^L = V_d^L$, the CKM matrix would be the identity matrix, and there would be no family mixing in weak-interaction processes. The existence of the CKM matrix is thus another example of the way in which the mass sector (through the Higgs mechanism) breaks the weak isospin symmetry. It also breaks nuclear isospin symmetry (the symmetry between up-type and down-type quarks), which acts symmetrically on left-handed and right-handed quarks.

Note that the mixing matrices V^R associated with the right-handed fermions do not enter into the Standard Model. They do, however, become relevant in extensions of the Standard Model, such as supersymmetric or left-right-symmetric models, and they can add to family-number violating processes.

Finally, we note that, because the neutrinos are assumed to be massless in the Standard Model, there is no mixing matrix for the leptons. In general, the leptonic analog to the CKM matrix has the form

$$V_{\text{lepton}} = V_\nu^L V_e^{L\dagger} \ .$$

But we are free to choose any basis for the neutrinos because they all have the same mass. By choosing the rotation matrix for the neutrinos to be the same as that for the charged leptons $V_\nu^L = V_e^L$, we have

$$\nu_0 = V_e^{L\dagger} \nu \ \text{ and } \ e_0 = V_e^{L\dagger} e \ .$$

The leptonic part of, for example, the charge-raising current is

$$\sum_i \nu_{0i}^\dagger \bar{\sigma}^\mu e_{0i} = \sum_{i,k,j} \nu_i^\dagger (V_e^L)_{ij} \bar{\sigma}^\mu (V_e^{L\dagger})_{jk} e_k = \sum_i \nu_i^\dagger \bar{\sigma}^\mu e_i \ ,$$

and the leptonic analog of the CKM matrix is the identity matrix. This choice of eigenstate would not be possible, however, if neutrinos have different masses. On the contrary, the neutrinos would have a well-defined mass eigenstate and there would likely be a leptonic CKM matrix different from the identity matrix. It is this leptonic mixing matrix that would be responsible for neutrino oscillations as well as for family-number violating processes such as $\mu \to e + \gamma$. ∎
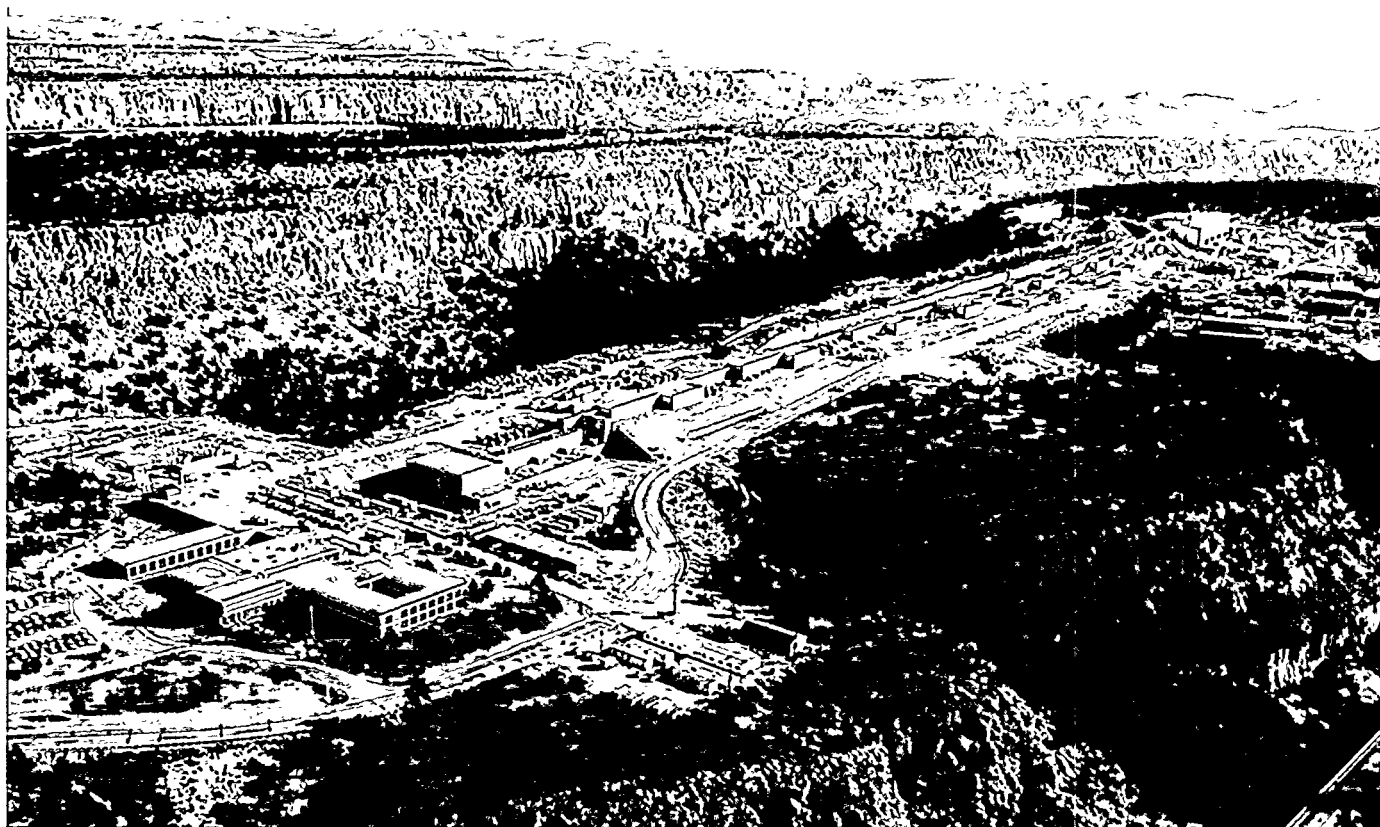
A Brief History of Neutrino Experiments at LAMPF

Gerry Garvey

Frederick Reines and Clyde Cowan Jr.'s observation of neutrino interactions in the late 1950s went largely unnoticed. It was overshadowed by the then recent, startling observation of parity violation in the weak interaction, an observation that flew in the face of cherished beliefs. Parity violation meant that the weak force had a handedness, a bias toward whether particles would spin right or left. In the case of the neutrino, nature always chose left. In 1932, when Wolfgang Pauli made the brilliant speculation that a nearly massless, neutrally charged particle must exist to explain perplexing features in nuclear beta decay, no physicists in their right mind would have suggested that such a particle also have the audacity to break left-right symmetry.

Parity violation evoked what is perhaps the most fundamental principle in science: the requirement to test, with ever more exacting experiments, the limits of prevailing theories and explanations. This arduous, challenging, and sometimes personally unrewarding search for the truth lies at the heart of the story of neutrino research. And nowhere is that story better exemplified than in the history of neutrino experiments at LAMPF (the Los Alamos Meson Physics Facility, now renamed the Los Alamos Neutron Science Center, or LANSCE).

The LAMPF accelerator came into operation in 1972 (see aerial photograph above and Figure 1). It was designed primarily to accelerate a high-intensity beam of protons to energies high enough to produce unbound pions.

Pions are short-lived, subatomic particles that are created when an energetic proton collides with a nucleus.

Neutrinos are a natural by-product of pion decay, and even before the accelerator was operational, physicists proposed exploiting that fact. Directing the unused portion of the beam into a large block of copper (called the beam stop) would produce pions that would come to rest within the beam stop. The positive pions, π^+, would decay into positive muons, μ^+, and muon neutrinos, ν_μ. (The negative pions would be reabsorbed by the copper nuclei before they decayed.) The positive muons would then decay into a positron, e^+, an electron neutrino, ν_e, and a muon antineutrino, $\bar{\nu}_\mu$. In all, three types of neutrinos would be produced—ν_e, ν_μ, and $\bar{\nu}_\mu$—that would radiate from
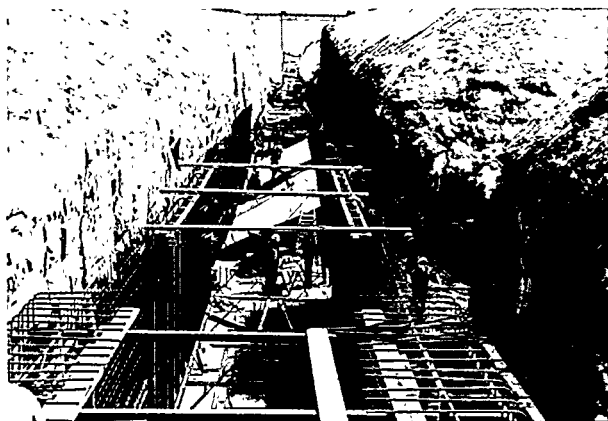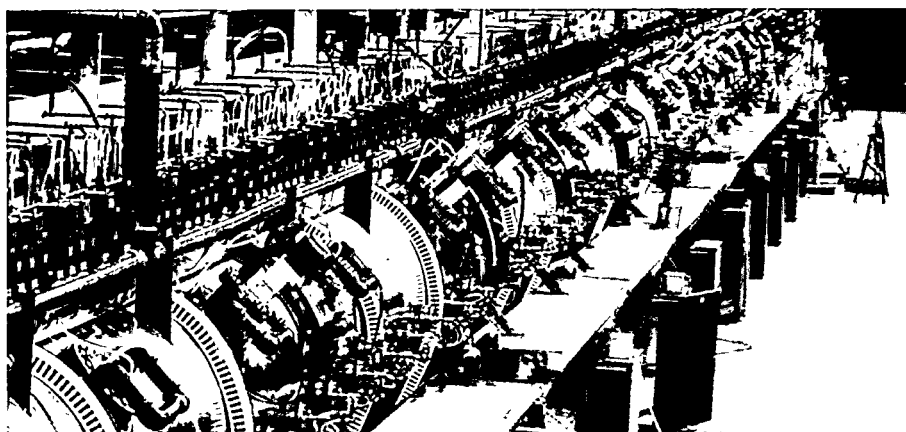
the beam stop in all directions.

The neutrinos produced would have energies between 10 and 55 million electron volts (MeV). In the early 1970s, neutrino interactions had been observed at only "low" energies (a few million electron volts) or "high" energies (roughly 1,000 MeV). Thus, LAMPF would enable the study of interactions at intermediate energies.

LAMPF had several unique properties that made it an almost ideal neutrino source. First, it had the highest instantaneous beam intensity of any of the existing, or proposed, meson factories (even though one never has "sufficient" intensity for neutrino experiments). In comparison with other high-energy accelerator sources, the intense LAMPF proton beam produced more neutrinos per second, so that one could anticipate more neutrino events in the detectors. Second, the average energy of the neutrinos was below the threshold for producing muons from muon neutrinos or muon antineutrinos.

Figure 1. A Brief Photo History of LAMPF
(a) Early photo of the trench dug into the mesa to accommodate the proton accelerator.
(b) Happy faces around the control console when the proton beam was first accelerated to design specifications (800 MeV). (c) LAMPF's first stage—an Alvarez linear accelerator—which brings the beam to an energy of 100 MeV. (d) Keyhole view of the accelerator's second stage, which brings the beam from 100 MeV to its final energy of 800 MeV. (e) LAMPF's end station, where experiments are carried out. The detector for the LSND experiment sits in the tunnel in the lower right-hand corner of the photo.

Table I. Neutrino Experiments at LAMPF

Experiment	Years	Reactions Observed	Principal Scientific Goals
E-31	1975–1980	$\bar{\nu}_e + p \rightarrow e^+ + n$ $\nu_e + D \rightarrow e^- + p + p$	Deduce the form of the muon-family-number conservation law
E-225	1975–1993	$\nu_e + e^- \rightarrow e^- + \nu_e$ $\nu_e + {}^{12}C \rightarrow e^- + X$	Measure the scattering cross section between electrons and electron neutrinos Measure the electron neutrino cross section on ${}^{12}C$ (X is another atom, typically ${}^{12}N$)
E-645	1980–1993	$\bar{\nu}_e + p \rightarrow e^+ + n$	Search for $\bar{\nu}_\mu \leftrightarrow \bar{\nu}_e$ oscillations
E-764	1982–1992	$\nu_e + {}^{12}C \rightarrow \mu^- + X$ $\nu_\mu + {}^{12}C \rightarrow \mu^- + X$	Search for $\nu_\mu \leftrightarrow \nu_e$ oscillations Measure the muon neutrino cross section on ${}^{12}C$
E-1173	1989–present	$\bar{\nu}_e + p \rightarrow e^+ + n$ $\nu_e + {}^{12}C \rightarrow \mu^- + X$	Search for $\bar{\nu}_\mu \leftrightarrow \bar{\nu}_e$ oscillations Search for $\nu_\mu \leftrightarrow \nu_e$ oscillations
E-1213	1990–present	$\nu_e + {}^{37}Cl \rightarrow e^- + {}^{37}Ar$ $\nu_e + {}^{127}I \rightarrow e^- + {}^{127}Xe$	Measure the cross section for electron neutrino capture on ${}^{37}Cl$ and ${}^{127}I$ to calibrate solar-neutrino detectors

This meant that electron neutrino interactions could be studied without interference from the muon neutrino processes that dominate experiments at higher energies. Third, because the proton beam was bunched in time, the neutrinos were only created during short intervals. An experimenter knew precisely when a neutrino produced by LAMPF could enter the detector. Events that occurred outside of those time windows would be the result of background processes.

There was one other feature of LAMPF that was favorable to neutrino experiments. Neutrinos were produced primarily from positive pion and muon decay. Aside from knowing very well the flux and energy spectrum of each neutrino type that was produced, experimenters also knew that electron antineutrinos were *not* produced. Therefore, an excess flux of electron antineutrinos in their experiment could be interpreted as evidence for neutrino oscillations.

All these advantages were outlined in a proposal that was written before

LAMPF began operation (Lande and Reines 1971). The proposal was prophetic insofar as it anticipated what was to be the LAMPF neutrino program for the next 20 years. It called for several specific experiments to be carried out when the proton beam neared its design intensity of 1 milliampere (equal to 6×10^{15} protons per second). Four experimental goals were outlined: (1) to deduce the form of the lepton-family-number laws, in particular, the electron- and muon-family-number conservation laws; (2) to measure the scattering cross section between electrons and electron neutrinos; (3) to measure the neutrino interaction cross sections that were relevant to solar-neutrino experiments; and (4) to search for neutrino oscillations.

Ken Lande and Fred Reines wrote the proposal, but they had input from many of the outstanding scientists of the day, including Clyde Cowan, Vernon Hughes, Hans Frauenfelder, Darragh Nagle, and Ray Davis. Also contributing were some of the younger

researchers who were later to provide much of the technical innovation and drive necessary to make the LAMPF neutrino program a success: Bob Burman, Herb Chen, Don Cochran, and Peter Nemethy. The proposed neutrino source was built, and Don Cochran and Lou Agnew assumed primary responsibility for its operation.

All told, six experiments have been conducted using the LAMPF neutrino source. A brief summary of them is given in Table I. The remainder of this article discusses these experiments, although the focus is on the three experiments that have had the most-far-reaching consequences. Each of the experiments was a sizable undertaking involving several institutions plus the resources and technical personnel at LAMPF. But it is equally important to note that each experiment, while executed to achieve its own goals, was also a precursor for the next. Experience gained from one experiment, like stepping stones, helped researchers to cross uncharted waters.

Experiment E-31

Headed by Vernon Hughes and Peter Nemethy, this experiment examined the manner in which the muon-family-number is conserved. It had been established in the early 1960s that a positive muon decayed by transforming into a positron and two neutrinos. With our current understanding of lepton families and the weak interactions, we would write the decay as

$$\mu^+ \rightarrow \bar{\nu}_\mu + \nu_e + e^+ . \qquad (1)$$

Muon decay is entirely analogous to the beta decay of the neutron. As written, Reaction (1) also obeys separate, additive lepton-family-number conservation laws.

A conservation law simply means that whatever is present at the start of a reaction is also present—in the same amount—at the end of the reaction. Separate additive conservation laws meant that for each lepton family (either the electron family or the muon family), the *sum* of the family numbers before and after a reaction would be the same. Table II lists the first two lepton families with their family numbers and demonstrates both additive and multi-plicative conservation laws. (See the primer, "The Oscillating Neutrino," on page 28 for a more detailed discussion of muon decay and the lepton-family-number conservation laws.)

In the early 1970s, many of the conservation laws, especially those involving the muon, still needed to be confirmed. Most physicists viewed the muon as a mysterious particle. It appeared to be simply a heavy version of the electron, and no one could understand why nature would summon up such a beast. The mathe-matical structure of the weak interactions was not well established, and there were no unbreakable laws governing muon decay.

Indeed, when E-31 was proposed in the early 1970s, all the available data were consistent with the four possible lepton-family-number conservation laws

Table II. Lepton-Family Numbers and Possible Conservation Laws

Lepton	Electron-Family Number, L_e	Muon-Family Number, L_μ
e^-	+1	0
ν_e	+1	0
e^+	−1	0
$\bar{\nu}_e$	−1	0
μ^-	0	+1
ν_μ	0	+1
μ^+	0	−1
$\bar{\nu}_\mu$	0	−1

Possible conservation laws:
1. Additive: $\Sigma\, L_e$ and $\Sigma\, L_\mu$ separately conserved
2. Multiplicative: $\Sigma\, (L_e + L_\mu)$ and $(-1)^{\Sigma\, L_e} (-1)^{\Sigma\, L_\mu}$ separately conserved
3. $\Sigma\, (L_e + 2L_\mu)$ conserved
4. $\Sigma\, (L_e - L_\mu)$ conserved

Reaction (1) in the text obeys separate additive conservation laws:

$$\mu^+ \rightarrow \bar{\nu}_\mu + \nu_e + e^+$$

$0 = 0 + 1 + (-1)$	Sum of electron-family numbers is conserved.
$-1 = -1 + 0 + 0$	Sum of muon-family numbers is conserved.

Reaction (1) also obeys the multiplicative law:

$[0 + (-1)] = [0 + 1 + (-1)] + [(-1) + 0 + 0]$	$\Sigma\, (L_e + L_\mu)$ is conserved.
$(-1)^0 (-1)^{-1} = (-1)^0 (-1)^{-1}$	$(-1)^{\Sigma\, L_e} (-1)^{\Sigma\, L_\mu}$ is conserved.

The reaction also obeys the third and fourth conservation laws.

listed in Table II. However, the *multiplicative* conservation law allowed a second muon decay channel:

$$\mu^+ \rightarrow e^+ + \bar{\nu}_e + \nu_\mu . \qquad (2)$$

Reaction (2) had never been observed. It was strictly forbidden by the much more theoretically appealing additive law. (The sum of the electron-family numbers is −2 after the reaction, instead of 0, so that the conservation law is violated. The sum of the muon-family numbers is also not conserved. Therefore, the reaction should not occur.) If the muon did decay by this mode, some of the guiding principles about the weak interactions would have to be reevaluated. It was of interest to see if muons decayed by this channel at all, and if so, to make an accurate measurement of the relative rates between Reactions (1) and (2).

LAMPF was ideally suited to perform such an experiment because the facility relied on positive muon

Figure 2. E-31 Collaborators and Detector
From left to right, Robert Burman, Donald Cochran, Jean Duclos, and Peter Nemethy stand in front of the E-31 detector, the first water-filled Cerenkov detector used to search for neutrinos.

decay to produce its neutrinos. An experiment could verify the forbidden decay mode simply by observing electron *antineutrinos* in a suitably built detector.

Hughes, Nemethy, and their collaborators used a water-filled Cerenkov detector to search for electron anti-neutrinos (see Figure 2). They looked for the signature of a positron emerging from the reaction

$$\bar{\nu}_e + p \rightarrow e^+ + n \; , \qquad (3)$$
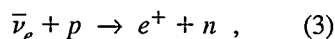
which is the same inverse-beta-decay reaction exploited by Reines and Cowan to observe the first neutrino interactions. But E-31 detected no electron antineutrino events coming from positive muon decay. Thus, the probability for Reaction (2) to occur had to be very small, below the sensitivity limits of the experiment. It appeared that the additive conservation law was correct to a very high level of accuracy and that family number was separately conserved by each lepton

family. (The later observance of a few electron antineutrinos in experiment E-1173 is now interpreted as evidence for oscillations of muon antineutrinos into electron antineutrinos. The possibility that Reaction (2) might still occur is described in the article "The Nature of Neutrinos in Muon Decay and Physics Beyond the Standard Model" on page 128.)

E-31 also carried out the only measurement of the cross section for electron neutrinos on deuterium, D. In order to calibrate the detector, the experimenters filled it with heavy water (D_2O) and observed the reaction

$$\nu_e + D \rightarrow e^- + p + p \; . \quad (4)$$
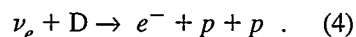
Comparing the frequency of events to the known electron neutrino flux yields the cross section. Reaction (4) is directly related to the primary energy-generating reaction in our sun:

$$p + p \rightarrow D + e^+ + \nu_e \; . \quad (5)$$
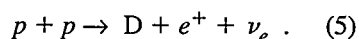
This is the *pp* reaction that has figured so prominently in the solar neutrino problem (see the article "Exorcising Ghosts" on page 136.)

Experiment E-225

Parity violation in nuclear beta decay was discovered in 1956 by Chien-Shiung Wu and her collaborators at the National Bureau of Standards. Shortly thereafter, Richard Feynman and Murray Gell-Mann formulated the $V-A$ theory (a "left-handed" theory that violated parity) for what is now called the charged-current weak interaction. The theory was immediately confirmed in a flurry of experimental and theoretical activity.

During the sixties and early seventies, powerful new theoretical insights by Sheldon Glashow, Abdus Salam, Steven Weinberg, George Zweig, and Gell-Mann, supplemented by numerous experimental observations at high-energy accelerators in the United States and Europe, led to a partial unification of the weak and electromagnetic interactions. The hallmark of the new electroweak theory was the inclusion of neutral-current interactions, which were mediated by the exchange of a neutral boson (Z^0). These neutral interactions were in addition to the well-studied charged-current interactions, which in the new electroweak theory were mediated by the exchange of a W^+ boson.

E-225 was proposed before neutral currents were discovered. Its original intent was to observe the charged-current scattering of electron neutrinos from electrons and to measure the cross section. In that reaction, the incoming electron neutrino transforms into an electron, and the target electron is transformed into an electron neutrino:

$$\nu_e + e^- \rightarrow e^- + \nu_e \; . \qquad (6)$$
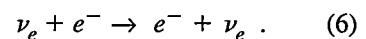
With the introduction of the electroweak theory, the objective of E-225 was quickly changed. In addition to the charged-current interaction, the new

theory predicted that Reaction (6) could also proceed through neutral-current scattering, in which both the electron neutrino and the electron maintained their identities as they scattered from one another. The two distinct modes of interaction meant that two terms entered into the calculation of the cross section and could potentially "interfere" with each other. The electroweak theory of Glashow, Salam, and Weinberg predicted a destructive interference, meaning that the cross section would be less than what was expected for just the charged-current scattering. The new objective of E-225 became to confirm or disprove that prediction. The experiment was headed by Herb Chen, a very talented young man who was in many ways the leader of the neutrino physics community at Los Alamos during this time. Unfortunately, Chen died of leukemia in 1987.

The experiment used a detector that was built like a 40-layer sandwich, with each layer made of plastic scintillator and flash-chamber module. A single module (see Figure 3) contained 10 flash-tube panels, with each panel containing 520 flash tubes. A flash tube is a long, narrow tube of gas that outputs a current pulse when a charged particle passes through it. A panel of 520 flash tubes could provide one-dimensional position information for a particle with very good resolution. Within a flash-chamber module, panels were alternately arranged either horizontally or vertically, so that each module could track a charged particle in two dimensions.

The stack of 40 modules (containing a total of 208,000 flash tubes) enabled the detector to provide a three-dimensional trajectory of an electron with very good position resolution. Experimenters knew that the scattered electron emerging from Reaction (6) would follow a trajectory that was confined to a narrow cone surrounding the neutrino's direction. Trajectory information, combined with energy information provided by the plastic scintillators, allowed the experimenters

Figure 3. A Flash-Chamber Module
Bobby Rechtor (center) prepares to lift a flash-chamber module (silver plane.) Forty modules made up the neutrino detector for the E-225 experiment. At left are Minh Van Duong and Robert Burman; at the back are Peter Doe and K. C. Wang. The man at the right is unidentified.

to identify those electrons that came from neutrino-scattering events.

E-225 found that the scattering cross section ruled out constructive interference between neutral- and charged-current interactions, and thus the experiment was altogether consistent with the predictions of the electroweak theory. It also confirmed the widely held belief that, when passing through electron-rich matter, electron neutrinos have a different scattering cross section than do muon or tau neutrinos. (The latter neutrino types can only interact with electrons through neutral-current scattering.) The different cross section was also applicable to the solar-neutrino problem. If neutrino oscillations do occur, then electron neutrinos born in the core of the Sun would scatter differently than would the neutrinos into which they oscillate. This is the fundamental assumption underlying the MSW effect, which is the most-favored solution to the solar neutrino problem. (See the article "MSW" on page 156.)

Experiments E-645, E-764, and E-1173

The Standard Model assumes that neutrinos are massless. Consequently, there can be no mixing between the three lepton families, and hence lepton-family numbers are separately conserved in every interaction. However, there appears to be no fundamental reason for a massless neutrino. Furthermore, any extension of the Standard Model that leads to neutrinos with mass also leads to mixing between the lepton families. Therefore, a neutrino that has mass will likely be a mixture of the three neutrino types and will have some probability to oscillate between them.

E-645 was undertaken in the early 1980s with the specific goal of searching for the oscillation of muon antineutrinos into electron antineutrinos. The experiment was headed by Tom Romanoski. Although it did not find any evidence for oscillations, for a time it established the upper limit on the

Figure 4. E-645 and the Cosmic-Ray Veto Shield

(a) E-645 began with the excavation of a tunnel to house the experiment. The structure in the lower left is the cosmic-ray veto shield in an early stage of construction. (b) An inside view of the black, archlike veto shield. The shield was movable and was rolled into the completed tunnel to check clearances. The inner set of railroad tracks allowed the E-645 detector to be rolled under the shield. (c) With the detector in place beneath the shield, the electronics "caboose" was joined to the end. The veto shield and detector were then pushed the remaining few meters into the tunnel.
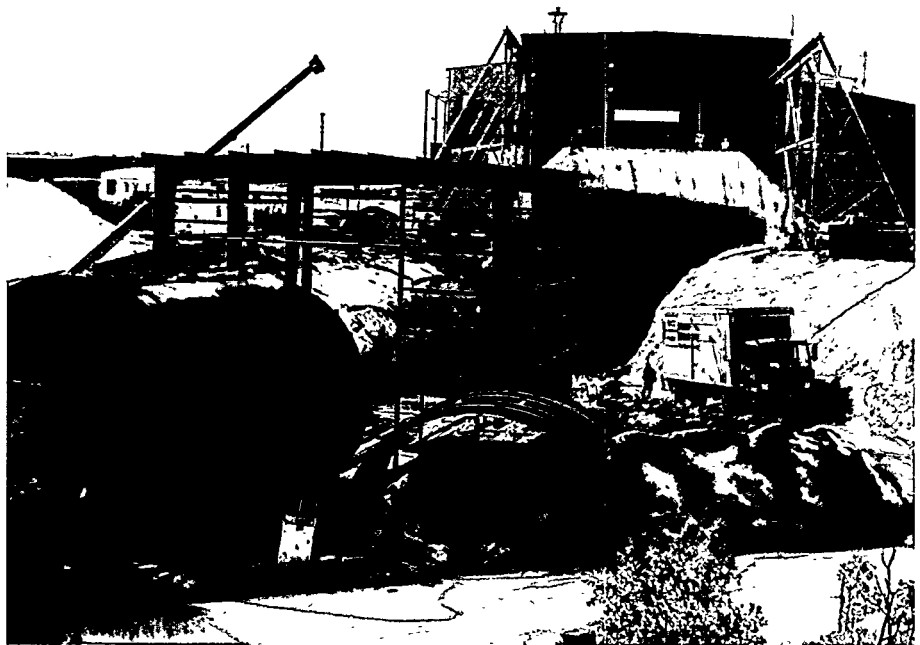
(a)
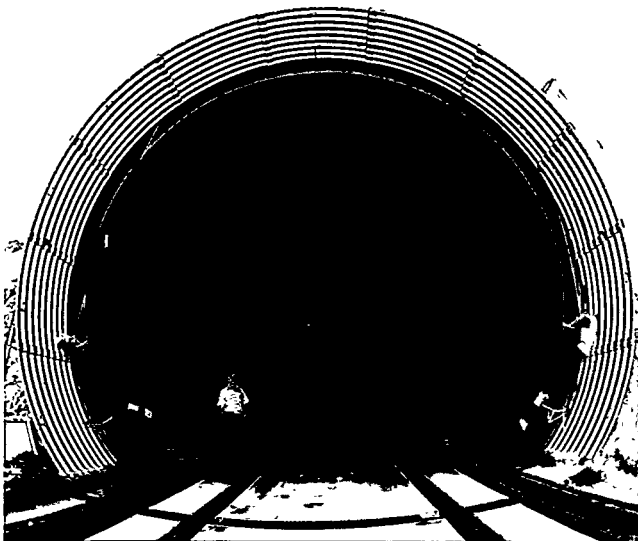
(b)
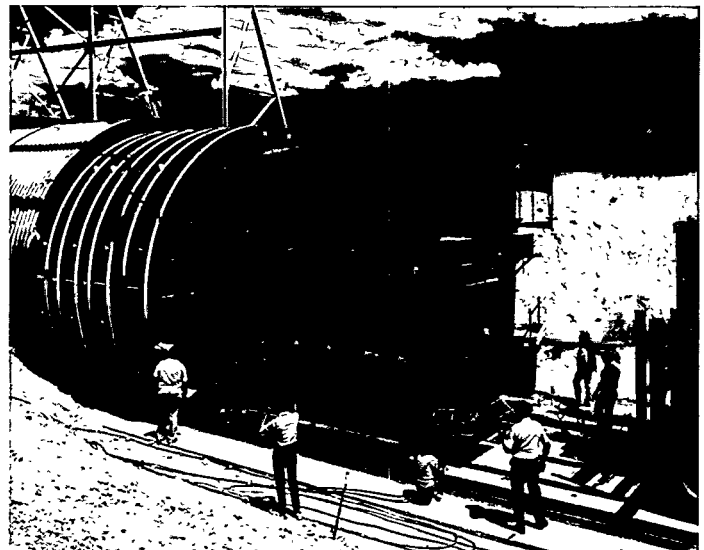
(c)

probability of oscillations taking place. The experiment also had long-term consequences in that it produced a piece of equipment known as the cosmic-ray veto shield. The double-walled shield weighed over a thousand tons and surrounded the bulk of the detector. It was filled with liquid scintillator and would send out a signal when a cosmic ray passed through, thus allowing the experimenters to reject a huge number of background events. The veto shield was a marvelous piece of equipment that was gratefully used by later neutrino experimenters (see Figure 4).

E-764 was to be a follow-up to E-645. Headed by Tom Dombeck, it investigated the use of the proton storage ring (PSR) as a low-duty-factor, decay-in-flight muon neutrino source. Unfortunately, the experiment was plagued with many difficulties, most notably a high background rate and a low initial neutrino flux (because the PSR was still being commissioned). As a result, E-764 was administratively terminated. Researchers were able to obtain a new upper limit for the oscillations, but mostly they gained a heightened awareness for how difficult it is to do neutrino experiments.

The oscillation experiment that is currently running, E-1173, also known as LSND for its liquid scintillator neutrino detector, represents a giant step forward in the neutrino program at LAMPF (see Figure 5). The detector is nearly 10 times larger than the ones used in E-225 and E-645. (The detector was designed to be as large as possible, constrained only by the need to fit inside the E-645 veto shield.) The detector has a trigger that is 5 times more efficient than any used by earlier experiments, and data is gathered 50 times faster. A new proton

Figure 5. Members of Experiment E-1173 (the LSND Collaboration)
The large tank in the background is the LSND detector, which is filled with mineral oil and scintillator.

beam target was even built with the purpose of providing higher-energy muon neutrinos that would enable a search for muon neutrino to electron neutrino oscillations.

The neutrino source and detector improvements have allowed LSND to detect a surplus of events ascribable to electron antineutrinos, which the experimenters believe provides evidence of oscillations. If this result is confirmed, the experiment will prove that neutrinos have mass and will provide the first experimental evidence for physics beyond the Standard Model. (See the article "A Thousand Eyes" on page 92 for more information on this experiment.)

Experiment E-1213

This ongoing experiment is trying to measure the capture cross sections for electron neutrinos on ^{37}Cl and ^{127}I. These elements are used as targets in detectors that are looking at solar neutrinos, such as Ray Davis' chlorine experiment in the Homestake Mine

in South Dakota and a new radiochemical experiment (IODINE), also installed in the Homestake Mine.

To extract the ^{127}I cross section, the E-1213 detector is filled with 1.5 tons ($\approx 7 \times 10^{27}$ atoms) of ^{127}I in the form of sodium iodide dissolved in water. "Interaction of the iodine nucleus with electron neutrinos creates ^{127}Xe, which is periodically extracted from the "detector. The analysis for the ^{127}I experiment is continuing.

The Legacy

Beginning with Reines and Cowan's experiments that were followed by more than 20 years of neutrino experiments at LAMPF, Los Alamos has a history of neutrino physics for which it can be truly proud. In many ways, the success of the neutrino physics program here at Los Alamos and at other national laboratories is a tribute to the creative vitality of these institutions, often maintained in the face of bureaucratic conservatism. The research begins with a burst of enthusiasm, high

hopes, and optimistic schedules. Unfortunately, the reality is that the experiments take a great deal of time, taxing both the resources and the patience of the experimenters, and the rewards, if any, often come only after the initial researchers have left.

But the human intellect is compelled to understand, rather than simply describe, nature's phenomena, and neutrino experiments have provided unique and crucial insights into the structure and processes of our physical universe. It is somewhat ironic that the nearly undetectable neutrino has had such an impact on scientific knowledge. ■

Further Reading
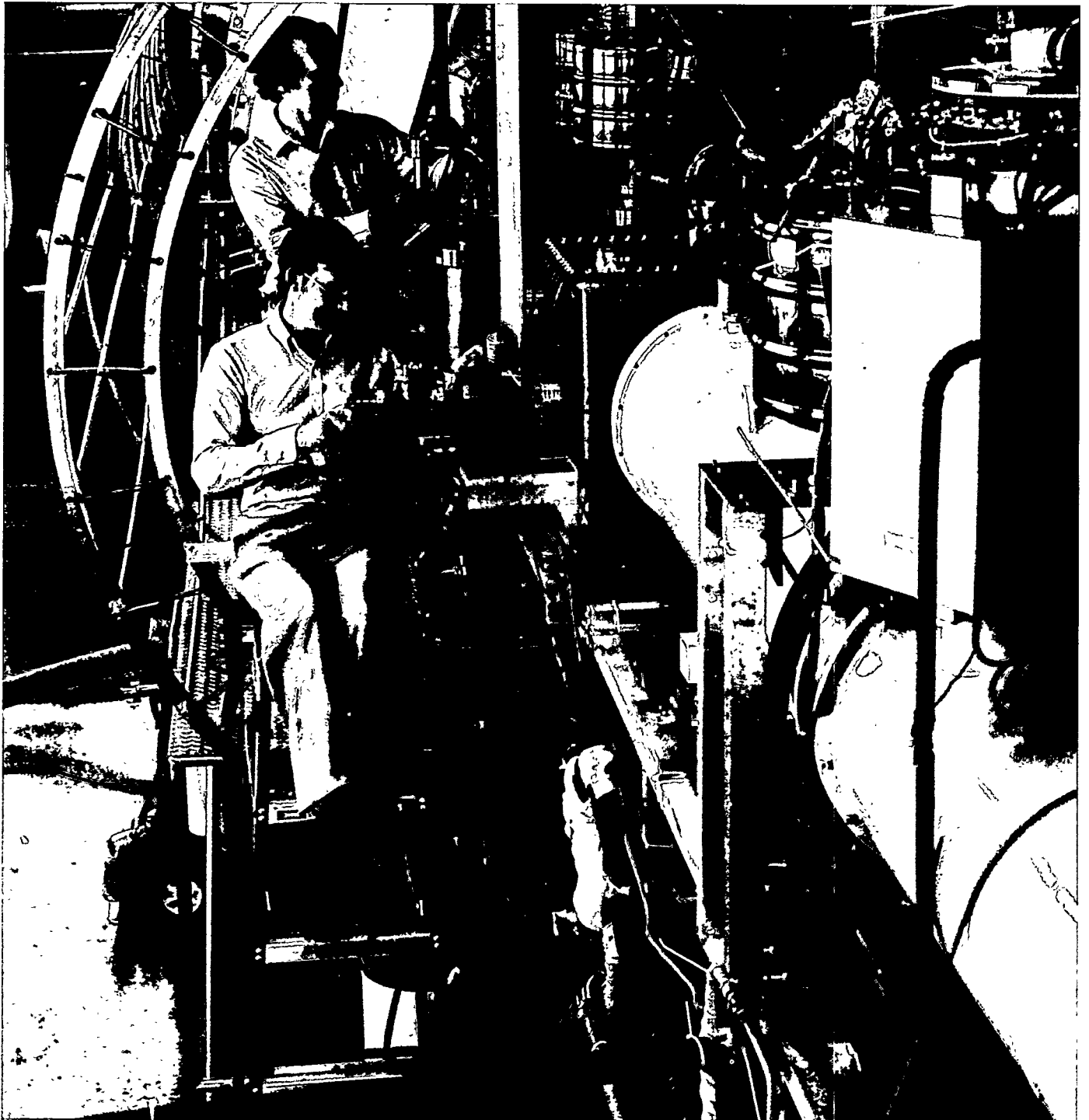
Lande, K., and F. Reines. 1971. LAMPF Neutrino Facility Proposal. Los Alamos Scientific Laboratory report LA-4842-MS. (Online at http://lib-www.lanl.gov/la-pubs/00362846.pdf)

Gerry Garvey is the former director of LAMPF. His biography appears on page 63.

Tritium Beta Decay and the Search for Neutrino Mass

Thomas J. Bowles and R. G. Hamish Robertson as told to David Kestenbaum

Neutrinos have been around, literally, since the beginning of time. In the sweltering moments following the Big Bang, neutrinos were among the first particles to emerge from the primordial sea. A minute later, the universe had cooled enough for protons and neutrons to bind together and form atomic nuclei. Ten or twenty billion years later—today—the universe still teems with these ancient neutrinos, which outnumber protons and neutrons by roughly a billion to one. Stars such as the sun churn out more; Wolfgang Pauli himself was unknowingly awash in trillions of solar neutrinos while he was drafting his "desperate remedy."[1]

We tend to think of neutrinos as transients, interacting only through the weak force and gravity and tracing long, lonely trajectories across the universe. But what they lack in strength they make up in number. Even if neutrinos were to have a mass as small as one billionth of that of a proton or neutron, their cumulative tug would be enormous, affecting the gravitational evolution of the universe as much as the normal matter we observe every day. It is believed that a neutrino mass of 22 electron volts would cause our universe to contract and eventually collapse because of gravitational forces.

Ironically, all who attempted to measure the mass of the neutrino directly used the very process that compelled Pauli to postulate its existence more than sixty years ago—the curious phenomenon of beta decay. Early experiments determined that certain radioactive atoms produced beta particles (high-energy electrons) when they decayed. The law of energy conservation dictates that the electron should emerge with a specific energy, identical every time, as it recoils against the atom. The electrons, however, appeared with a variety of energies, and Pauli correctly inferred

that the decay also produced a second unseen particle, now called the electron neutrino. The neutrino would share the energy released in the decay with the daughter atom and the electron. The electrons would emerge with a spectrum of energies.

In 1934, Enrico Fermi pointed out that, if the neutrino had mass, it would subtly distort the tail of this spectrum. When an atom undergoes beta decay, it produces a specific amount of available energy that is carried away by the electron, the neutrino, and the daughter atom. Typically, the bulky atom remains relatively still, while the electron and neutrino split the available energy. Sometimes, the electron takes more than half, sometimes less. On extremely rare occasions, it can carry off nearly all the energy.

This maximum amount of energy the electron can carry off is called the endpoint energy and marks the tail end of the spectrum of electron energy released in the decay. If the neutrino has no mass, the endpoint energy is very nearly equal to the energy released in the decay. On the other hand, Fermi pointed out, a finite neutrino mass would make the end-point energy slightly lower and shorten the tail of the spectrum.

If some of the energy released in the decay were "locked up" in the mass of the neutrino, it would be unavailable to the electron, and the mass of the neutrino could be determined from a careful measurement of the spectrum near the endpoint. Unfortunately, the converse (a massless neutrino) can never be proved; it is always possible that the neutrino has a small mass that lies just beyond the reach of the latest experiment. A Zen-like axiom underlies this quandary: you cannot weigh something that has no mass.

The ideal beta-decay source has a short lifetime and releases only a small amount of energy in the decay. A small energy release means that more decays fall near the endpoint, where the shape of the electron energy spectrum is sensitive to a small

neutrino mass. A short lifetime means atoms decay more rapidly, making more data available.

A wonderful accident of nature, tritium (a hydrogen atom with two extra neutrons) is a perfect source by both of these measures: it has a reasonably short lifetime (12.4 years) and releases only 18.6 kilo-electron-volts (keV) as it decays into helium-3. Additionally, its molecular structure is simple enough that the energy spectrum of the decay electrons can be calculated with confidence.

The predicted spectrum (shown in Figure 1) peaks at around 4 keV and extends up to the endpoint energy, around 18.6 keV. Only one out of every 10 million decays emits an electron in the last 100 electron volts before the endpoint, where the shape is sensitive to neutrino masses in the range of 30 electron volts (see close-up of the endpoint), so testing the tail requires precision as well as patience.

ITEP Weighs in with Neutrino Mass

Was the neutrino mass holding back some energy from the electron? In 1980, the answer seemed to be a startling "yes." Over the years, numerous experiments had probed the endpoint with increasing precision and concluded that the neutrino could have a mass no more than a few tens of electron volts. But in 1980, Russian scientists at the Institute for Theoretical and Experimental Physics (ITEP) in Moscow announced that they had pushed even further and discovered a shortfall near the endpoint corresponding to a neutrino mass of around 35 electron volts. The consequences of such a hefty mass would be enormous. The Standard Model would have to be revised, and the universe would eventually collapse, albeit not for another 40 billion years or so.

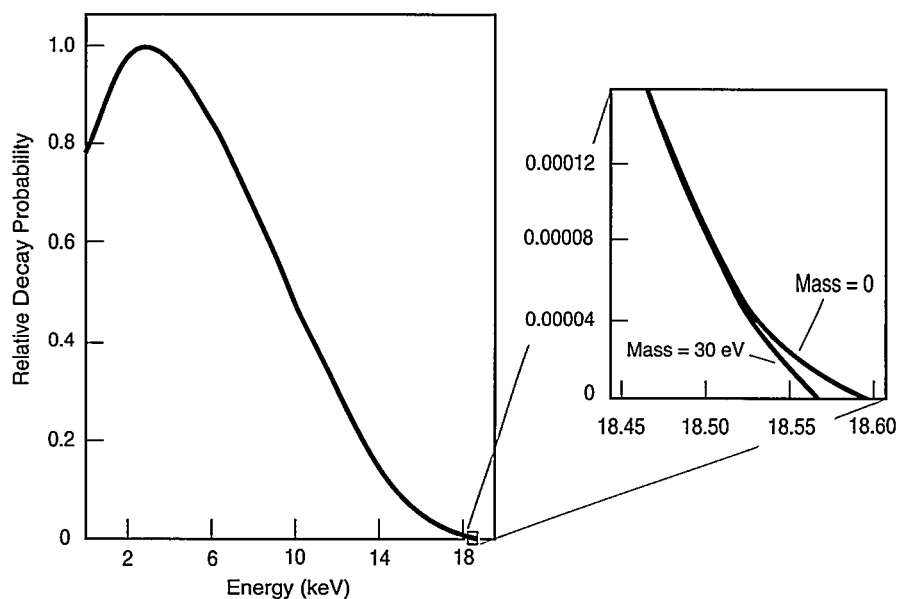But were the results correct? Investigations uncovered problems in the

Figure 1. The Beta Decay Spectrum for Molecular Tritium

The plot on the left shows the probability that the emerging electron has a particular energy. If the electron were neutral, the spectrum would peak at higher energy and would be centered roughly on that peak. But because the electron is negatively charged, the positively charged nucleus exerts a drag on it, pulling the peak to a lower energy and generating a lopsided spectrum. A close-up of the endpoint (plot on the right) shows the subtle difference between the expected spectra for a massless neutrino and for a neutrino with a mass of 30 electron volts.

calculation of the spectrum shape and errors resulting from the energy resolution of the Russian spectrometer. Members of the ITEP group carefully and methodically conducted a new round of experiments checking for these and other systematic errors and providing new data. Although they reduced their prediction of the electron neutrino mass to 26 electron volts, their original conclusion that the neutrino has mass remained the same.

Still, there were many ways to generate a slump at the end of the spectrum and mimic a finite neutrino mass: the electrons could be losing energy from scattering off other atoms in the source, the spectrometer resolution could be off, or some energy could be tied up in an unanticipated excited state of the daughter atom. In particular, the community voiced concern over ITEP's use of a solid source, an amino acid called valine in which some of the hydrogen atoms had been replaced with

tritium. Valine was convenient because it was readily available, but its complex molecular structure meant that the atoms were left in a myriad of excited states following the decay. The excitations could rob the electron of energy and, if not properly taken into account, could induce an apparent erosion of the spectrum near the endpoint. Moreover, the excitation energies were quite similar to the observed neutrino mass, and a difficult and uncertain theoretical calculation was needed to correct for the effect. Thus, the ITEP claim left room for considerable doubt.

The Los Alamos Experiment: Simple in Theory, Tough in Practice

Several months before the ITEP announcement, over gelati at the Erice conference in Italy, Los Alamos physicists Thomas Bowles and

Hamish Robertson (now at the University of Washington) had decided they would also join the hunt for a neutrino mass. With the salvo that the ITEP measurement drew, there was no better time to enter the fray.

In 1980, armed with innovative methods designed to circumvent the uncertainties that had cast doubt on the earlier work, a team at Los Alamos led by Robertson and Bowles began an exhaustive search for the electron neutrino mass. Instead of a solid source, pure, gaseous, molecular tritium was used (see Figure 2). Molecular tritium (a bound state of two tritium atoms) was simple enough that theoretical physicists could accurately calculate the atomic excitation energies, taking into account all the interactions between the two electrons and nuclei. Even with this seemingly simple system, the calculations were involved, requiring many days of computation on a Cray computer. By contrast, the ITEP source, valine, contained 19 atoms and 64 electrons, making such a calculation intractable.

The use of a gas also reduced energy loss in the material and eliminated "backscattering" where the electron could hit the backing (used to support the solid source) and do an energy-sapping U-turn, which could produce a dip in the spectrum near the endpoint. But this theoretical simplicity came at the expense of experimental complexity. Handling a kilocurie of tritium gas posed many challenges. The complex arrangement of magnets, pumps, and other equipment for the experiment filled a room 30 feet by 70 feet.

But the grand contraption had a relatively simple task: To capture electrons from the beta decay of the tritium gas and carefully transport them to a high-precision magnetic spectrometer. Only those electrons that enter with a certain fixed energy can traverse the magnetic fields set up in the spectrometer. A silicon detector sits at the end of the spectrometer and counts the electrons that make it through.

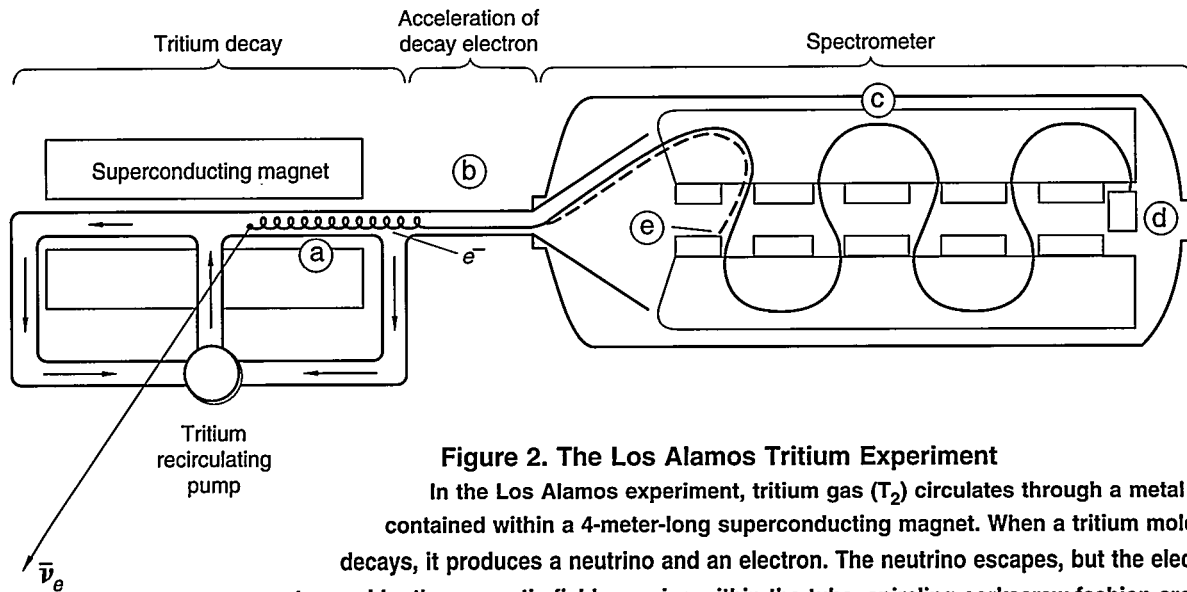The tritium gas that begins the whole process is circulated and

Figure 2. The Los Alamos Tritium Experiment

In the Los Alamos experiment, tritium gas (T_2) circulates through a metal tube contained within a 4-meter-long superconducting magnet. When a tritium molecule decays, it produces a neutrino and an electron. The neutrino escapes, but the electron, trapped by the magnetic field, remains within the tube, spiraling corkscrew-fashion around the field lines (a). The electron emerges from the magnet and receives a kick in energy (b) before it is passed to the spectrometer (c). Magnetic fields in the spectrometer guide the electron through several S-turns and focus them onto the detector (d). The magnetic fields are chosen so that only the electrons with energies near the endpoint reach the detector. Electrons with too little energy quickly get off course and run aground in the walls of the spectrometer (e).

recirculated through a long metal tube, 4 centimeters in diameter, which itself is contained in a larger-diameter solenoidal superconducting magnet. The magnetic field points along the axis of the tube, and it contains and guides the decay electrons without altering their energy. The electron neutrino, of course, leaves the tube, the room, and eventually the solar system, but the electrons remain, spiraling corkscrew-fashion in very tight, millimeter-radius circles along the field lines. The field strength varies along the tube so that the electrons are corralled toward one end of the 4-meter magnet. Electrons that head off toward the wrong end are bounced back by an increasing field gradient. When the electrons exit the magnet, a second magnetic field separates them from the gas before they are finally injected into the large toroidal spectrometer. Electrons near the endpoint energy have a velocity roughly one million meters per second, and their dizzying journey takes only a fraction of a second.

One concern was that tritium would accumulate inside the spectrometer.

Electrons resulting from its decay could bypass the difficult obstacle course and pollute the data with spurious "background" counts. The Los Alamos group solved this problem by setting the spectrometer to count electrons of 23 or 24 keV (above the endpoint) and placing the tritium source at a higher voltage than the spectrometer's. The added voltage gave the electrons that entered the spectrometer an extra "kick" in energy. The silicon detector, in addition to counting the arriving electrons, was also designed to provide a rough measurement of the electron energy (accurate to about 3.5 keV) and, so, could be used to discriminate between the electrons coming from the source and the lower-energy ones coming from the tritium lodged in the spectrometer.

Transporting and measuring the electrons were delicate affairs, and care also had to be taken to eliminate any stray magnetic fields that could derail the electrons. An additional coil outside the spectrometer eliminated the earth's magnetic field.

Steel girders in the building had to be demagnetized by hand.

Another concern was that contaminants such as oxygen and nitrogen, which inevitably leak into the system, could build up. These atoms, which are relatively bulky compared with tritium, could skim off energy from the electron through inadvertent collisions. Forcing gas through a palladium filter removed the larger atoms and cleaned the system.

The tritium itself also presented a few sticky problems. Because tritium and hydrogen are effectively siblings (both contain one proton), the two often trade places, and the tritium ends up affixed to all manner of surfaces. Over time, for instance, tritium accumulates in the walls of the tube, taking the place of hydrogen atoms that used to be there. To ensure that the electrons reaching the spectrometer originated from the gas and not the tube walls, the physicists tuned the spectrometer to accept only electrons that came from the very center of the tube. This had the unfortunate consequence of

stripping away 90 percent of the electrons from decays in the gas, but successfully reduced the number of electrons coming from the walls of the tube by a factor of 100,000 or more. Building an instrument is one thing; understanding what it does it quite another. Taking data with an uncalibrated device is like playing an out-of-tune piano. The result is more noise than music. In this case, the tuning had to be very precise: the energy measurements good to nearly one part per thousand. Fortunately, there was an elegant way to test the response of the apparatus—simply replacing the tritium gas with gaseous krypton-83m (an isotope of krypton that produces monoenergetic electrons). Krypton-83m is another wonderful accident of nature. It produces electrons close in energy (17.8 keV) to the tritium endpoint, and so it is perfect for calibrating the spectrometer.

Each of the numerous tritium atoms circulating through the system had, every second, a one-in-a-billion chance of decaying. Roughly, sixty-million electrons of all energies entered the spectrometer every minute, of which only one, on average, had an energy near the endpoint that would carry it through the selective fields of the spectrometer. What began as a flood of electrons was reduced to a trickle of only one every minute. The physicists could only drum their fingers and wait for the drops to accumulate.

Seven Years Later: A Verdict and a New Mystery

In 1987, the Los Alamos scientists had finished an initial measurement and, by 1991, they had a clear verdict: the measurement of the tritium beta-decay spectrum showed no deficit near the endpoint. This finding was consistent with an electron neutrino mass of zero and notably inconsistent with ITEP's results. A very tiny mass might have escaped detection, but it could not have been larger than 9.4 electron volts, which is
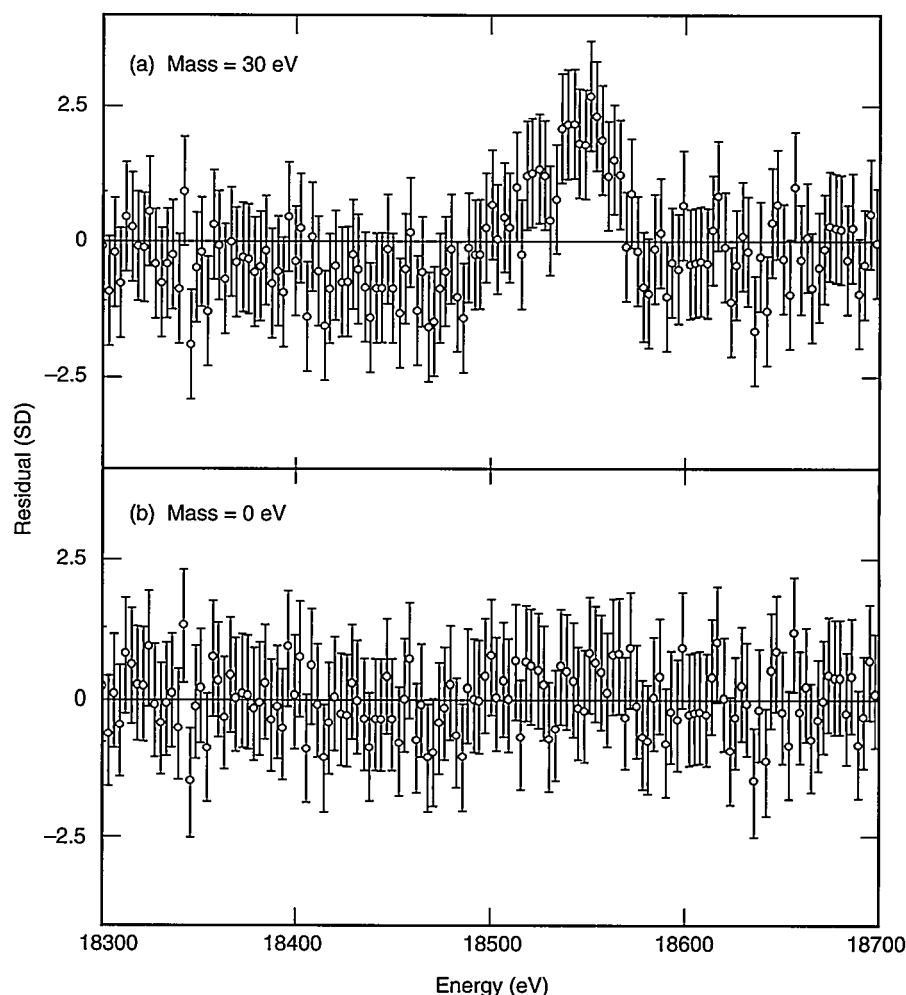
Figure 3. Did the Neutrino Weigh 30 Electron Volts?
Not according to the Los Alamos data. The top figure shows the data points from the tail of the spectrum compared with the expected values (the straight line) for an electron neutrino with a mass of 30 electron volts. The data wander from the line, ruling out the possibility of a 30-electron-volt neutrino. On the other hand, the bottom figure shows the same data points compared with the expectation for a neutrino mass of zero. While the data clearly favor a neutrino mass of zero (the points lie close to the line) over a mass of 30 electron volts, the best fit is actually for a slightly *negative* neutrino mass. (Note that in the bottom plot, the data points lie, on average, slightly above the line, so this is not a perfect fit.) Both plots display "residuals," which indicate how many standard deviations each data point is from a particular hypothesis. One can think of plotting the data over the top of the predicted spectra shapes of Figure 1, pulling the tail out so that it lies horizontal, and adjusting each data point so that its distance to the line is represented in standard deviations. (Each point has an experimental uncertainty associated with it. Two-thirds of the time, the true value is expected to lie within plus or minus one "sigma" or standard deviation from the point.)

far smaller than the 22 electron volts needed to cause the universe to contract. Figure 3 shows the data compared with the expected shape for a neutrino mass

of 30 electron volts and for a neutrino mass of zero.

But from the ashes of the Russian result arose a new mystery. Careful

inspection of the Los Alamos data revealed a small, curious surplus near the endpoint. A deficit would have meant that neutrinos had mass (see Figure 1), but a surplus did not make any sense. Although unlikely (the odds were roughly 1 in 30), the result could have simply been a statistical fluctuation.

Over the years, several other experiments have also ruled out the Russian result and confirmed the strange surplus near the endpoint (Stoeffl and Decman 1995 and Weinheimer et al. 1993). The surplus can no longer be explained away as a statistical fluctuation, and it prevents experimenters from establishing a tight upper limit on the neutrino mass. As stated in the Review of Particle Physics, the accepted encyclopedia of particle properties, "Given the status of the tritium results, we find no clear way to set a meaningful limit on m_{ν_e}."

Today, the tritium quandary has spawned a small cottage industry of professional speculators. There are, possibly, as many theories to explain the surplus as there are groups investigating it. The exotic possibilities run from tachyonic (traveling faster than the speed of light) neutrinos, to a new force that would cause clumping of neutrinos around our galaxy. More mundane explanations include unanticipated molecular or atomic effects in the tritium decay. Still, the simple structure of molecular tritium is thought to be well understood, and the calculations that yield the shape of the spectrum rest solidly on the time-proven laws of quantum mechanics.

It may be that what began as a search for neutrino mass has unearthed something far stranger. Experiments designed to ferret out whatever is hiding in the tail are on the drawing boards, but given the enormous technical challenges involved, headway will be hard won. Neutrinos had been around for billions of years before Pauli noticed them, and it may be a few more before their true character is revealed. ■

Thomas J. Bowles received his Ph.D. degree in 1978 from Princeton University. After a postdoctoral appointment at Argonne National Laboratory, he joined the Physics Division of the Laboratory in 1979. Bowles initiated a program in weak-interaction physics in the Physics Division, working on problems in beta decay, neutrino studies at LAMPF, and nuclear astrophysics. This program was initially centered on measurements of the tritium beta-decay spectrum as a sensitive means of searching for a finite mass of the electron antineutrino. The Los Alamos experiment was the first to employ a windowless free-molecular-tritium source. The results from this experiment refuted the claims of a Russian group who claimed to have measured a finite neutrino mass. They also ruled out electron antineutrinos as a possible candidate for most of the dark matter of the universe. Subsequently, Bowles became involved in studies of solar neutrinos as a means to extend the experimental sensitivity to a finite mass of the neutrino. In 1986, Bowles became the U.S. principal investigator on the Russian-American Gallium Experiment and a member of the Sudbury Neutrino Observatory project. Most recently, he initiated a program to develop a source of ultracold neutrons at LANSCE in order to study fundamental symmetries of nature in neutron beta decay. Bowles was elected Fellow of the American Physical Society in 1992, Los Alamos National Laboratory Fellow in 1994, and was appointed as an affiliate professor at the University of Washington in 1995.

R. G. Hamish Robertson was born in Ottawa, Canada, and was educated in Canada and England. He earned his undergraduate degree at Oxford University and his Ph.D. degree in atomic-beam and nuclear-structure physics at McMaster University in 1971. Robertson went to Michigan State University as a postdoctoral fellow and remained on the faculty, becoming professor of physics in 1981. In 1976, he received an Alfred P. Sloan Fellowship, and his research resulted in the first observation of an isobaric quintet of states in nuclei. Additionally, he carried out experiments on parity violation, nuclear astrophysics, and nuclear reactions. In 1981, he joined Los Alamos National Laboratory, becoming a Fellow in 1988, and was responsible for investigating neutrino mass via tritium beta decay and solar-neutrino physics. In 1994, Robertson took a professorship at the University of Washington, where he continued his work in neutrino physics. In 1997, he received the American Physical Society (APS) Tom W. Bonner Prize. He is a member of the Canadian Association of Physicists, an associate member of the Institute of Physics (London), and a Fellow of the APS. Robertson has chaired the Nuclear Science Advisory Committee and served on its Instrumentation Subcommittee. He is a member of the Board of Physics and Astronomy of the National Research Council (NRC) and has served on the NRC's Nuclear Physics and Neutrino Astrophysics Panels and the APS Division of Nuclear Physics Executive and Program Committees. Robertson has served on review committees for the Lawrence Berkeley National Laboratory's Nuclear Science Division and Caltech's Physics, Mathematics, and Astronomy Division, the Editorial Board of Physical Review D, and review panels for the National Science Foundation and the Department of Energy.

Further Reading

Belesev, A. I., et al. 1995. Results of the Troitsk Experiment on the Search for the Electron Antineutrino Rest Mass in Tritium Beta-Decay. *Physics Letters B* **350**: 263.

Lubimov, V. A., et al. 1980. An Estimate of the Upsilon-E Mass from the Beta-Spectrum of Tritium in the Valine Molecule. *Physics Letters* **94**: 266.
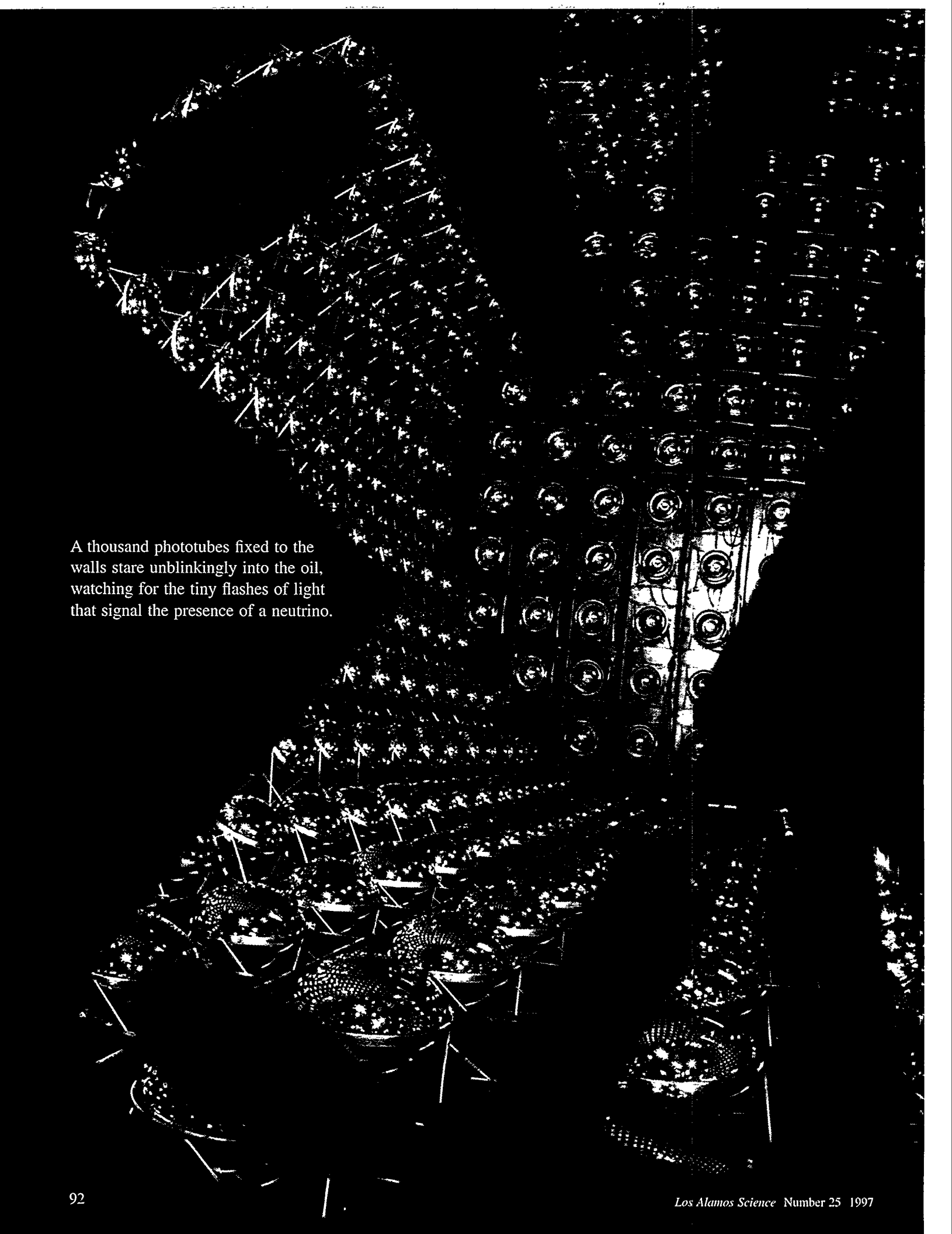
Robertson, R. G. H., et al. 1991. Limit on Electron Antineutrino Mass from Observation of the [Beta] Decay of Molecular Tritium. *Physical Review Letters* **67**: 957.

Barnett, R. M., et al. 1996. Review of Particle Physics. *Physical Review D* **54**: 1.
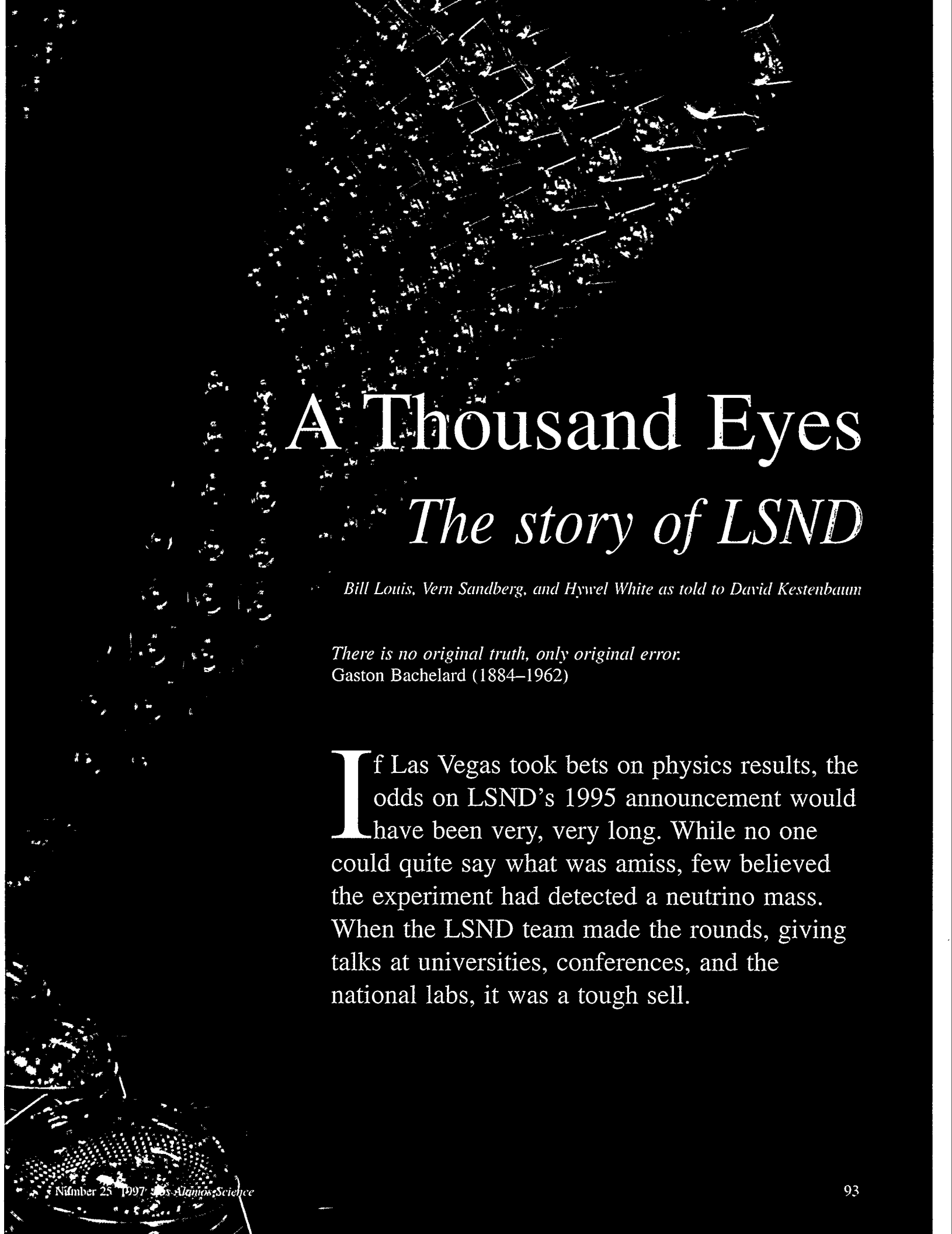
Stoeffl, W., and D. J. Decman. 1995. Anomalous Structure in the Beta Decay of Gaseous Molecular Tritium. *Physical Review Letters* **75**: 3237.

Weinheimer, C., et al. 1993. Improved Limit on the Electron-Antineutrino Rest Mass for Tritium Beta-Decay. *Physics Letters B* **300** (3): 210.

Wilkerson, J. F., et al. 1987. Limit on Electron Antineutrino Mass from Free-Molecular-Tritium Beta-Decay. *Physical Review Letters* **58**: 2023.

A thousand phototubes fixed to the walls stare unblinkingly into the oil, watching for the tiny flashes of light that signal the presence of a neutrino.

A Thousand Eyes

The story of LSND

Bill Louis, Vern Sandberg, and Hywel White as told to David Kestenbaum

There is no original truth, only original error.
Gaston Bachelard (1884–1962)

If Las Vegas took bets on physics results, the odds on LSND's 1995 announcement would have been very, very long. While no one could quite say what was amiss, few believed the experiment had detected a neutrino mass. When the LSND team made the rounds, giving talks at universities, conferences, and the national labs, it was a tough sell.

"The community was . . . intrigued, to say the least," recalls Ion Stancu, an LSND (liquid scintillator neutrino detector) collaborator from the University of California, Riverside. Stancu gave one of the first talks that January to a room packed past fire codes. The question period ran longer than the talk itself. "Some thought it was complete rubbish, others were very excited," he says. "In the end, all I could say was 'this is the data . . . take it or leave it.'" Many preferred to leave it, thinking it would just go away. Previous experience suggested it might.

For one, the Standard Model of particle physics states unequivocally that neutrinos are massless, and the Standard Model had yet to be proved wrong. Although the notion that neutrinos might have mass was not new—measurements of atmospheric- and solar-neutrino rates pointed to a similar conclusion—the LSND result didn't coincide with many physicists' expectations. Most theoretical models had to be stretched quite a bit to accommodate all three sets of data; there just didn't seem to be room for yet another positive result. (See the article "The Evidence for Oscillations" on page 116.)

Pressed to pick the wrong results from the lineup, many in the field suspected LSND's. Rather than wait for neutrinos from the heavens, LSND manufactured its own with a kilometer-long particle accelerator. In principle, this afforded greater control over the experiment. In practice, it had been a lesson in humility. Similar experiments had checkered histories—their claims seemed to flit in and out of existence like the neutrinos themselves.

Outliving this legacy would be hard, and rumors that several LSND collaborators were questioning the results did not help. Neither did the fact that the results had appeared on the front page of the *New York Times* before they were made public to the physics community, all at a time when the experiment was under the budget axe. But what, then, to make of the results? The LSND team argued the odds were only

1 in 1,000 that their results were wrong. Still, no one was placing bets.

LSND—A Walking Tour

Los Alamos, with its long history of defense work, may seem an odd place for the delicate task of weighing the neutrino. Above ground, in the foothills of the Jemez Mountains, there is little to betray the intricate machinery. The accelerator lies buried under a kilometer-long mound of dirt. Seen from an airplane, it recalls the inhumanly large constructions of the ancient Mayas,

designed to catch the eyes of the gods. At its far end, metal blocks, planks, and bricks are stacked several meters high like the abandoned toys of a giant toddler. Most are recycled relics from the cold war—iron from magnets at Oak Ridge Lab, steel from chopped-up battleships, and counterweights from missile silo doors. In their retirement, they shield a giant underground neutrino detector from cosmic rays and an occasional rattlesnake seeking refuge in the cracks.

The detector is so well shielded that it is all but impossible to get to. In a small shack nearby, a ladder leads
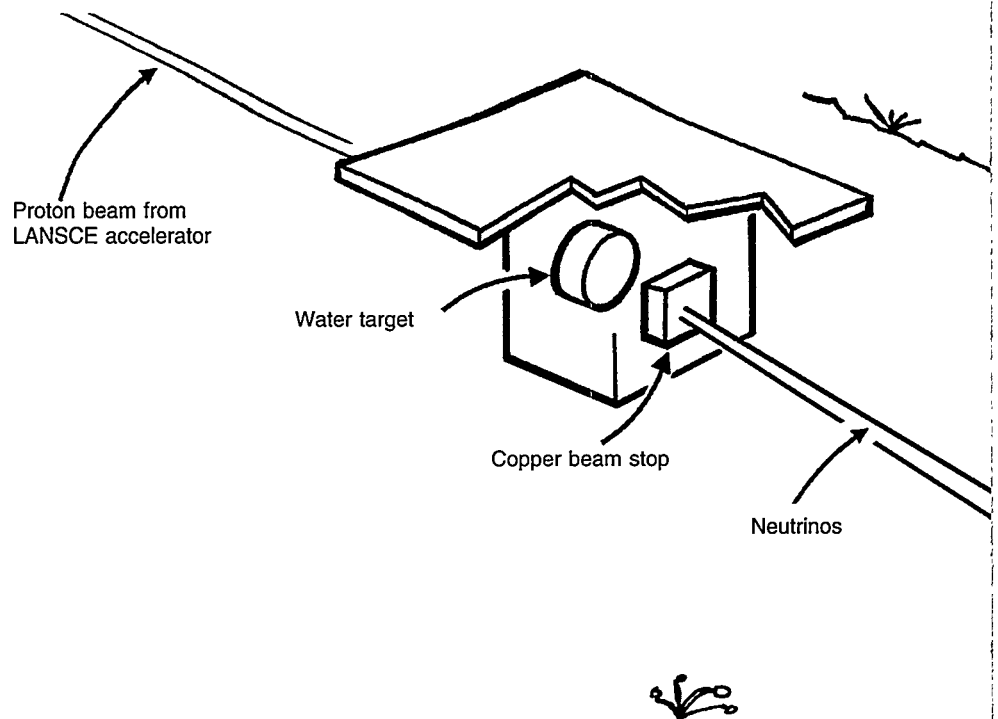
Figure 1. LANSCE Neutrino Source and LSND Detector
Neutrino production at LANSCE begins with pions, which are created when a high-energy proton beam (from the LANSCE accelerator) strikes a water target. The pions leave the target, travel a short distance in air, and most come to rest in a copper beam stop. There they decay into muons and neutrinos, and the muons decay into positrons and more neutrinos. The neutrinos fly off in all directions. Passing right through tens of meters of dirt and concrete, some of the neutrinos enter the LSND detector: a large, cylindrical tank containing over a hundred tons of mineral oil and studded on the inside with more than a thousand unblinking, electronic eyes. The tank sits on a thick floor of steel planking, while more steel shielding surrounds its front and top. Salvaged from battleships and other cold war relics, the shielding helps protect the tank from cosmic rays and rogue beam particles. A large plug of water also shields the detector, as does the archlike cosmic-ray veto shield that envelopes the tank. The veto shield clues experimenters when something other than a neutrino enters the tank.

down a narrow metal pipe into the dry rocky ground. Climb down, crawl through another sewer-size pipe, flip the light switch, and you'll find yourself in a small, dusty room fondly known as the Black Hole, silent except for the whir of fans cooling the electronics. The 6-meter-diameter circle of steel that forms one wall of the room is yet more shielding—the front face of a monstrous archlike shell called the cosmic-ray veto shield. Like the shielding blocks, the veto shield is a relic, this time from a previous neutrino experiment. Nestled inside it is the main detector: an enormous tank filled with 52,000 gallons of mineral oil. This is the heart of LSND, where over a thousand phototubes fixed to

the walls of the tank stare unblinkingly into the oil, watching for the tiny flashes of light that signal the presence of a neutrino. If a phototube fails, it's left for dead. Once the tank has been stuffed into the underground tunnel, only neutrinos and cosmic rays can get inside.

Fortunately, after years of troubleshooting by humans, the detector can essentially take care of itself. Most of the time, it clicks happily away, analyzing the electronic pulses from the phototubes with an elaborate array of hardware and then writing the data to magnetic tapes half the size of a cigarette pack. When a tape is full, the detector swaps it for a new one. If something goes drastically wrong, it pages a physicist for help. And if, somewhere among the millions of cosmic rays, it senses the pattern of lights that could signal a neutrino, it writes the information to a computer disk so that the forty LSND collaborators can log-in from their distant desks at universities coast to coast and check on the day's catch.

Figure 1 shows a sketch of the underground accelerator and neutrino detector.

How to Weigh a Neutrino

If neutrinos have mass, it is so slight that it hardly impedes their motion. Were it possible to produce a neutrino that stood perfectly still, the tiniest tap would suffice to send it fleeing to the ends of the universe at, or close to, the speed of light. And because the neutrino is electrically neutral, it cannot be grasped with electric or magnetic fields the way electrons or protons can. The only possibility of detecting neutrinos at all is through the weak force, which is roughly one hundred million times feebler than the electromagnetic force.[1] The weak force is the agent behind all neutrino behavior—how they produce flashes of light in the tank, how they are made, and even, perhaps, how they are "weighed."

Since it is impossible to sit a neutrino on a scale or to determine its mass by running it through the magnetic fields of a spectrometer, the neutrino can only be weighed indirectly. LSND, like atmospheric and solar experiments, looks for neutrinos to "oscillate," a strange behavior that can betray their mass.

Neutrinos come in three varieties—the electron neutrino, the muon neutrino, and the tau neutrino. Each neutrino also has an antimatter counterpart, called the electron, muon, and tau antineutrinos. When neutrinos (or antineutrinos) oscillate, they undergo a kind of identity crisis. An electron neutrino made in the Sun, for instance, may transform enroute to the earth and present itself instead as a muon neutrino or a tau neutrino. The probability of observing one neutrino type or another varies periodically as the neutrino travels, hence the term oscillation. Oscillations can occur only if neutrinos have mass. Definitive observation of neutrino oscillations would settle the

Black Hole area

Shielding blocks (salvaged relics of the cold war)

Cosmic-ray veto shield.

Steel

A thousand-eyed detector

Water plug (more shielding)

---
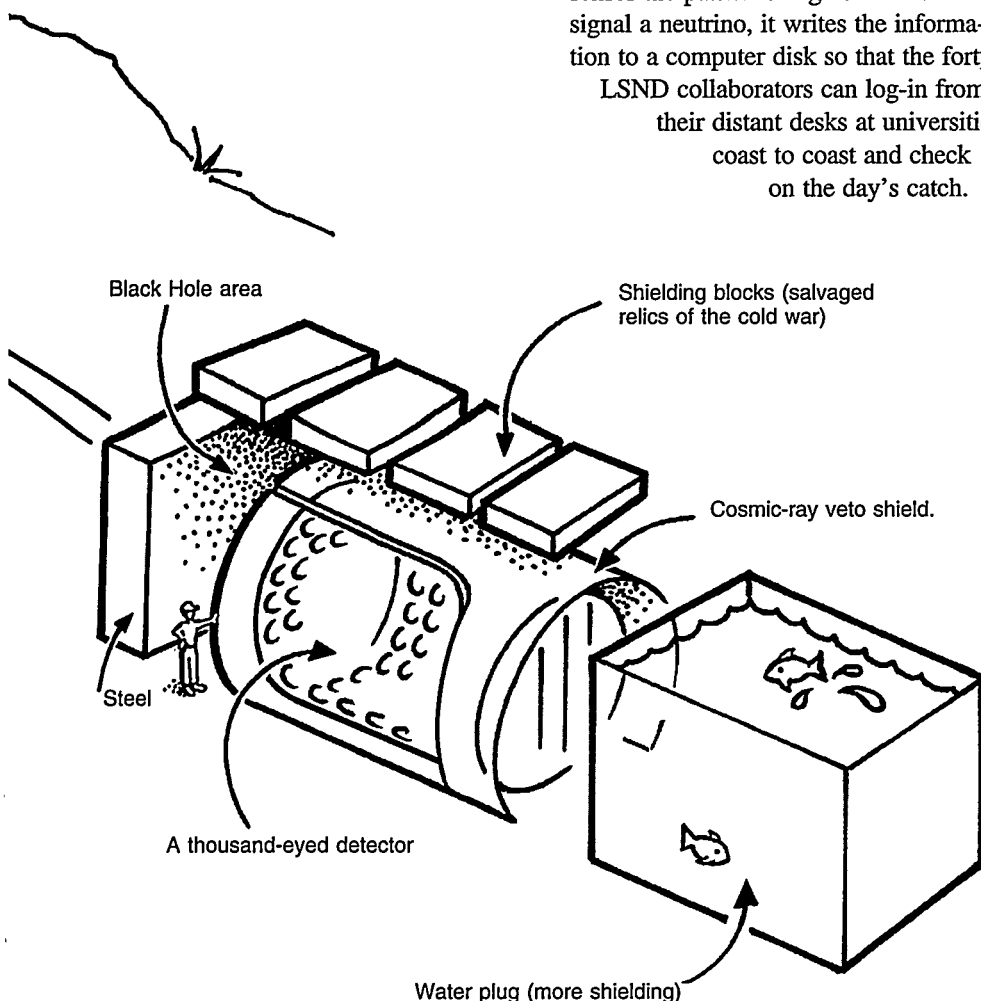
[1]Neutrinos feel gravity's tug, but too weakly to be of use experimentally.

Figure 2. A Neutrino Oscillation Experiment

If neutrinos oscillate, the probability of observing a given neutrino type varies with distance. An idealized oscillation experiment would consist of a neutrino source that produces only one type of neutrino, here muon antineutrinos (blue) from pion decay. Enroute to the detector, the muon antineutrino transforms into an electron antineutrino (red), which can interact with a proton in the detector to create a positron and a neutron. These two particles are taken to be the signature of the electron antineutrino.
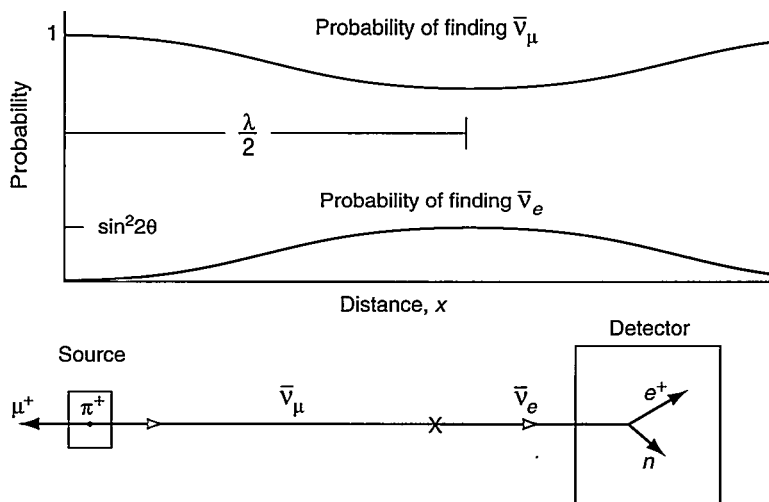
decades-old question of whether "massless" neutrinos are really massless.

Assume for the moment that neutrinos have mass, and furthermore, each neutrino would have a specific mass. The electron neutrino would weigh some amount m_1, the muon neutrino would weigh m_2, and the tau neutrino would weigh m_3—simple, elegant, and easy to explain to students. But in the paradoxical realm of quantum mechanics, a neutrino can be several things at once. In all likelihood, each neutrino has a split personality and possesses, in a sense, three masses. Put an electron neutrino on a scale, and it might read m_1, m_2, or m_3.

The three masses are like three ghostly neutrinos (called "mass states") that inhabit the electron, muon, and tau neutrinos. Mathematically, they can be seen as three ingredients which, combined in various proportions, form the electron, muon, and tau neutrinos.

To see how oscillations occur, imagine a muon neutrino produced in the decay of a muon. The muon neutrino can be viewed as a mixture of the three ghostly neutrinos, each with a different mass. The particular mixture (the recipe) that defines the muon neutrino is dictated by three numbers called "mixing angles." If other subatomic particles are any guide (that is, the quarks), one "ghost" will dominate each neutrino type so that, for

instance, a muon neutrino might be 90 percent m_1, 9 percent m_2, and 1 percent m_3 (see the article "The Oscillating Neutrino" on page 28).

The muon neutrino may oscillate into another neutrino type, because the mix of its ingredients can change as it travels. In the quantum picture, the neutrino is described by a "wave function" that can be seen as the sum of three separate waves, one for each of the mass states. As the neutrino travels, each mass-state wave "vibrates" at a frequency that depends on the neutrino's mass, so that the neutrino is like a three-note chord with each note beating against the others. The relative amounts of each mass state change with the rise and fall of one wave against another until the neutrino arrives at some detector designed to measure its type. Depending on how the three waves are synchronized at the detection point, the particle will have some probability of appearing as an electron, muon, or tau neutrino.

The mathematics of oscillations can fit on a single page (see the box "Derivation of Neutrino Oscillations" on page 52), but the trickier problem of why there is a primordial mixup of masses remains unsolved. The related problem of why particles have the masses they do also presents a conundrum that only a few broad-minded theorists have dared to tackle. Were

physics a religion, mass would have its own creation myth.

Blueprints

All neutrino oscillation experiments follow the same conceptual blueprints (see Figure 2). At the level of a sketch one might make on a napkin, there are only two components: a "source", which like a pitching machine, hurls out neutrinos of a known type, and a "detector", which like a catcher's mitt, absorbs and counts the neutrinos. The game is simple. If neutrinos have mass, they can oscillate as they travel, changing their identity back and forth as they go. Any difference between what the source throws out and what the detector observes can be chalked up to oscillations.

But most neutrinos fly straight through the detector, so oscillation experiments, like baseball, are a kind of long-attention span sport, with extended periods of thumb twiddling between bits of action. The slow pace is a reflection of the fact that the neutrinos must interact through the weak force. Because the weak force is so feeble, fewer than one in one trillion neutrinos will leave a mark in the tank. Even with LSND's high-intensity source, it is often an hour between neutrino catches.

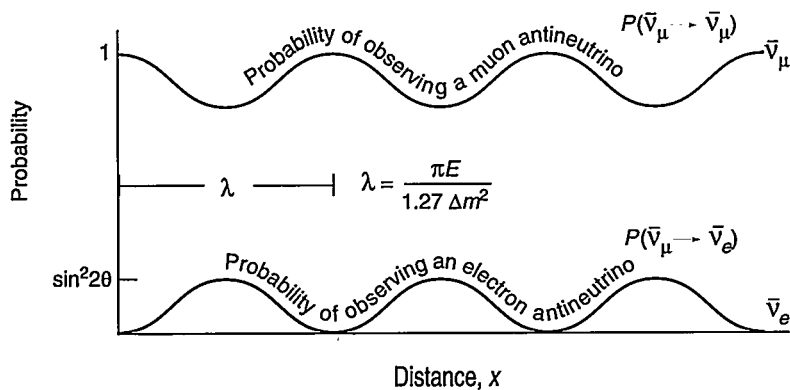Still, the weak force runs the show. Like any other force, it can push and

Figure 3. The Rise and Fall of the Electron Antineutrino

The probability of finding a muon antineutrino (blue curve) goes down as the probability of finding an electron antineutrino (red curve) goes up. Both probabilities oscillate with the same wavelength λ, and the sum of the two is always equal to one.

pull, but it can also create and absorb particles. In the source, the weak force gives birth to the neutrinos (here muon antineutrinos produced through the decay of a muon), and at the other end, it is again responsible for their demise.

The death is swift, if rare. The neutrino disappears, replaced by new particles that generate telltale patterns of light in the oil. The "disappearance" is really just a change in identity. The weak force tugs on the neutrino and the protons in the tank, and can occasionally "transfer" the electric charge of one proton to the neutrino. In the process, the proton becomes an (electrically uncharged) neutron, and the electron antineutrino becomes a positively charged electron (called a positron).[2] Both the neutron and the positron generate flashes of light in the oil which draw the attention of a roomful of electronics.

Thus the weak force anchors the experiment at both ends, creating and then providing the means of detecting the neutrinos. But when the neutrino flies between the source and the detector, the weak force is, in a sense, left behind. In the interim, the ghostly hand of quantum mechanics takes over and starts the neutrino oscillating.

[2]The neutrino names, in fact, derive from this pairing. The weak force transforms an electron neutrino into an electron, a muon neutrino into a muon, and a tau neutrino into a tau.

Feeling Around in the Dark

At the heart of every oscillation experiment lies a single all-important equation that gives the probability that a neutrino beginning its journey as one type will be observed as another type. For simplicity, assume that the tau neutrino is out of the picture and that oscillations take place only between a muon antineutrino and an electron antineutrino. (LSND searches for the latter particle.) In that case, the probability for a muon antineutrino to transform into an electron antineutrino is

$$P\left(\bar{\nu}_\mu \rightarrow \bar{\nu}_e\right) = \sin^2 2\theta \, \sin^2\!\left(\frac{1.27\Delta m^2 x}{E}\right) .$$

Here θ is the mixing angle (the "recipe" for how to combine two mass states to make a muon antineutrino), Δm^2 is the difference between the squares of the masses of the two mass states (that is, $\Delta m^2 = \left| m_2^2 - m_1^2 \right|$) and is in units of electron volts squared (eV²), E is the neutrino energy in million electron volts (MeV), and x is the distance between creation and detection in meters. Despite its complicated appearance, all the equation really means is that the probability of observing an electron antineutrino goes up and down like a sine wave as the distance to the source or the energy of the neutrinos is changed. (See Figure 3.) Indeed, since the mixing angle is a constant of the

world (albeit unknown), and since the energy of neutrinos is typically fixed for a particular experiment, the oscillation probability is often written as just

$$P\left(\bar{\nu}_\mu \rightarrow \bar{\nu}_e\right) = A \, \sin^2\!\left(\frac{\pi x}{\lambda}\right) ,$$

where the oscillation "wavelength" is given by $\lambda = (\pi E)/(1.27\Delta m^2)$ and the size of the oscillation is $A = \sin^2 2\theta$.

Given the difficulty of detecting neutrinos, every experimenter would love to place the detector where the probability of seeing an electron antineutrino peaks. The first place would be one-half of a wavelength from the source (see Figure 2). Unfortunately no one knows exactly how long a wavelength is.

Ideally, theoretical calculations could predict the neutrino masses that determine the wavelength, but so far such predictions lie beyond the scope of even the most far-reaching theories. From a theoretical perspective, searching for oscillations is like digging for buried treasure without a map. The oscillation wavelength could be 2 millimeters or 2 light-years.

The quest is made still more difficult because an experiment essentially only measures one number (here the number of electron antineutrinos) but seeks information about two quantities (Δm^2 and $\sin^2 2\theta$). If, after a year, an experiment saw nothing, it could mean that the mixing angle is very small (and hence A, which functions like a "volume" control, squelches the probability of observing an electron antineutrino) or it could mean that the detector happened to sit at a distance where the oscillation probability was low (a distance very much smaller than the wavelength for example). Then, the best one can do is to rule out the values of $\sin^2 2\theta$ or choices of Δm^2 that would have given an observable number of electron antineutrinos. Figure 4 shows the values that had been searched and ruled out by experiments before LSND began taking data in 1993.

Similarly, if after a year, the catcher gazes into the glove and miraculously sees a few electron antineutrinos, it is

impossible to sort out how much of the oscillation was due to A and how much to Δm^2. In fact, it wouldn't even be clear that oscillations had produced the electron antineutrinos. The appearance of electron antineutrinos could also be interpreted, perhaps more interestingly, as evidence for a new, bizarre decay of the muon, forbidden by the accepted laws of physics.[3]

This was the strange limbo that beset LSND following the 1995 announcement. Although LSND results showed the appearance of electron antineutrinos in a flood of muon antineutrinos, the measurement had been made at only one distance. Without seeing the number of detected electron antineutrinos rise and fall periodically as a function of the distance x (or as a function of the energy E), few were willing to write massive neutrinos into the textbooks.

The LSND collaborators themselves agreed that one point did not make an oscillation. To address this , LSND had been designed to see oscillations in a second way, by looking for the transformation of muon neutrinos into electron neutrinos, the "matter" counterpart of its primary antimatter analysis. This second analysis would later provide an invaluable cross-check on LSND results, but in 1995 it was not yet complete.

Like all neutrino oscillation experiments, LSND was a shot in the dark. If the experiment had indeed observed oscillations, it would have been a lucky happenstance that the source-to-detector distance was right.

Or of course, it could have been a mistake. The best-designed detectors are imperfect and can be duped by electronic noise, by other particles from the beam, or by the unrelenting rain of cosmic rays. Understanding these "backgrounds" formed the linchpin for the LSND experiment and was the focus of
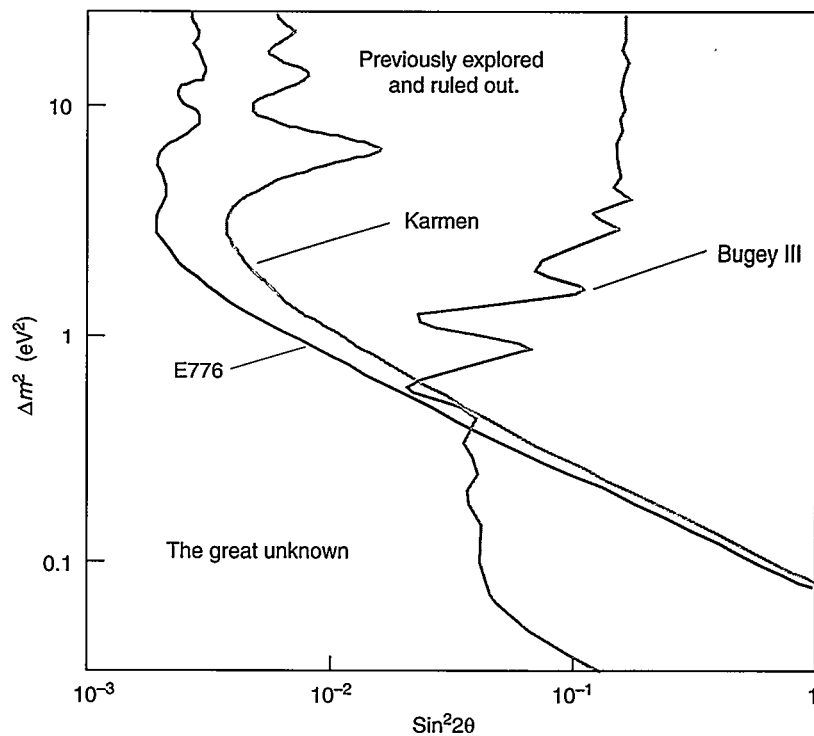
Figure 4. Values of Δm^2 and $Sin^2 2\theta$ Explored Before LSND
Several experiments had searched for neutrino oscillations and reported negative results. The values of Δm^2 and $sin^2 2\theta$ probed and ruled out are shown in grey: E776 (red) was an accelerator-based experiment at Brookhaven National Laboratory; Karmen (blue), an accelerator-based experiment at Rutherford Appleton Laboratory in England; and Bugey III (green), a French reactor-based experiment. When LSND began taking data in 1993, it was hoped that the experiment would have enough sensitivity to probe "the great unknown" region that remained.

the questions that filled the air whenever LSND researchers presented their results. While a seminar audience could not judge in an hour what had taken years to put together, many feared that the experimenters, too, had somehow been duped.

Celestial vs Terrestrial Neutrinos

LSND is the fifth in a series of neutrino experiments at Los Alamos, and its design draws heavily on the experience of its predecessors. Most importantly, it has inherited a decade-old high-intensity neutrino source based on the Los Alamos Neutron Science Center, or LANSCE, accelerator. An old warhorse, LANSCE is still the highest-intensity proton accelerator in

the world for its energy. Physicists are as familiar with its behavior as a soloist is with a well-rehearsed piece of music.

Earthbound neutrino sources like LANSCE outshine their celestial counterparts in many respects. From their far-flung birthplaces, celestial neutrinos must travel a long way to visit terrestrial detectors—thousands of light-years from supernovae, a hundred million kilometers from the Sun. As messengers, they are invaluable, bringing news of distant events such as the death of stars and reactions in the heart of the Sun; but as sources for oscillation experiments, their remote and uncertain origins make them less than ideal. There is no user's manual for the Sun that states exactly what kind of neutrinos are being produced, no dials to turn to adjust its output, and no switch to turn it off. Particle accelerators provide a way

---

[3]If some positively charged muons made by the accelerator decayed into electron antineutrinos, it would (falsely) appear as if oscillations had occurred. See the article "The Nature of Neutrinos in Muon Decay and Physics Beyond the Standard Model" on page 128.

Figure 5. LANSCE Production of Decay-at-Rest (DAR) Neutrinos and Their Energy Spectra

$$\pi^+ \rightarrow \mu^+ + \nu_\mu$$

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu$$

(a)

(a) LANSCE accelerates protons and shoots them into a water target, where they produce pions. Most of the pions travel a half meter in air before striking a copper beam stop and coming to rest. Nearly all of the negative pions (π^-) are absorbed by copper nuclei before they decay, but each positive pion (π^+) decays at rest to a muon (μ^+) and a muon neutrino (ν_μ). The muons also come to rest in the beam stop and decay to a muon antineutrino ($\bar{\nu}_\mu$), a positron (e^+), and an electron antineutrino ($\bar{\nu}_e$). The neutrinos fly off in all directions as from a miniature sun. (b) Energy spectra for the muon neutrinos (purple), muon antineutrinos (blue), and electron neutrinos (red) created in the beam stop. Although electron antineutrinos from negative pion decay have energies similar to those of electron neutrinos, these antineutrinos are produced so infrequently that their spectrum would appear as only a small dot on the plot.

to construct an intense neutrino source closer to home, a kind of sun-on-earth that can be tuned and tested.

But even a custom-made neutrino factory will have flaws. The ideal source would produce only one neutrino type so that oscillations could be easily identified. Neglecting the tau neutrino and its antiparticle still leaves four other types (electron neutrinos, electron antineutrinos, muon neutrinos, and muon antineutrinos) and, unfortunately, LANSCE makes them all. The goal, then, is to get rid of one (electron antineutrinos in this case) to clear a channel so that oscillations can be detected. The trick is preventive medicine—to stop electron antineutrinos before they are made. Miraculously, all this takes is a block of copper.

Neutrino production begins with a burst of protons from the kilometer-

long LANSCE accelerator. As seen in Figure 5, the protons strike a water target, producing pions. Virtually all pions will produce both muon neutrinos and antineutrinos when they decay. The positively charged pions will also make electron neutrinos, while the negatively charged pions will make electron antineutrinos.

The pions fly through the air and plow into a copper block, called the beam stop, where they slow and come to rest. Fortunately, the negative pions, because of their charge, are absorbed by the positively charged copper nuclei before they can decay, so that only a very few electron antineutrinos are produced. The positive pions, however, hang around until they decay into muons, which in turn decay to produce abundant neutrinos. Thus the copper block filters out the negative pions and

keeps the source relatively free of electron antineutrinos. In the end, the other neutrino types produced outnumber electron antineutrinos by a factor of roughly 10,000 to 1.

It should be noted that a small number of pions decay in flight before reaching the copper block. These pions produce higher-energy neutrinos than the pions that decay at rest in the copper block. The "decay-in-flight" (DIF) neutrinos thus become a second, separate source riding piggyback on the first and can be used to cross-check the results from the decay-at-rest (DAR) neutrinos. In fact, the DIF neutrinos are the subject of the second analysis alluded to earlier.

LANSCE loses out to neutrino sources like the Sun, however, in one important category. The problem isn't physical, it's fiscal. The sun shines for

free, but LANSCE eats up about $20 million in electric bills every four months. LANSCE serves several experiments in addition to LSND, but in 1993, the Department of Energy (DOE) threatened to shut down the power-hungry accelerator in order to feed other research efforts.

At the time, LSND researchers were just preparing to take their first data. If LANSCE died, LSND would follow, buried in its underground crypt. At the last minute LANSCE was spared, and the researchers, scrambling to finish LSND construction, took a total of six weeks of data during September and October. Even in this small amount of data, there were signs that a few extra electron antineutrinos had reared their heads. The researchers found eight electron antineutrino-like events[4] when they had expected only 0.9 ± 0.2.

The results, which were published in a conference proceedings the next year (Louis et al. 1994), drew considerable interest, but more data was needed to confirm the excess of events. If people knew one thing about neutrino physics, recalls Richard Imlay, "it was that it was hard. Many people wanted to wait and see if the results held up." Despite the cloudy funding picture, LSND was cleared to take another three and a half months of data beginning in August of 1994.

At the time, *Scientific American* devoted a page to the LSND story. "Missing Matter Found?" it asked. Not yet. "We feel we have a high burden of proof," it quoted Hywel White, one of the Los Alamos collaborators, as saying, because if neutrinos have mass, "it's very important." After the upcoming LSND run, the article quipped, "the team at Los Alamos should be able to verify—or otherwise—their nonclaim" (Mukerjee 1994).

---

[4]This was a faint heartbeat. If correct, it implied that the oscillations were small indeed, since LSND would have seen 5000 events if all the muon antineutrinos produced had fully oscillated into electron antineutrinos.

Detecting the Electron Antineutrino

Neutrinos produced in the LANSCE beam stop travel at velocities near the speed of light through steel shielding, earth, and concrete, finally reaching the detector 30 meters downstream. Over this short distance, the muon antineutrinos may change their identity and oscillate into electron antineutrinos. If the weak force, then, should chance to connect an antineutrino to a proton in the detector's oil, a single positron and neutron will emerge, each producing light as it passes through the oil.

The mineral oil that fills the tank is composed of carbon and hydrogen (chains of 30 or so CH_2 molecules). The electron antineutrinos are detected when they react with the hydrogen atoms (which are essentially free protons) through a process called "inverse beta decay,"

$$\bar{\nu}_e + p \to n + e^+ ,$$
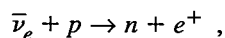
so named because it represents the reverse of the normal "beta decay" process common in radioactive nuclei.

The creation of the neutron and positron heralds an electron antineutrino event. Both particles lead to the production of light in the detector, and it is by observing that light that LSND knows an event has taken place. By human standards the light is invisibly faint, but to the 1,220 phototubes that can see a single photon and measure its arrival time to a nanosecond, the light shines like a miniature pyrotechnic display.

The positron generates light through two mechanisms. When created, the positron has a velocity greater than the speed of light in the oil and produces an electromagnetic shock wave analogous to the wake of a speedboat or to the sonic shock wave of a Concorde breaking the sound barrier. The light, called Cerenkov radiation, forms a cone that expands along the positron's trajectory like the headlight of a tiny car. The cone has a 47-degree opening angle and forms a shell rather than a

solid. Projected onto a flat surface, the cone leaves a telltale ring.

The positron also produces light with the help of a small amount of scintillator that is added to the oil. As the positron travels, it loses energy by inducing atomic excitations in the oil; by a secondary process, the scintillator also gets excited. When the scintillator de-excites, it produces blue light that is detected by the phototubes. The excitation process takes a little time and delays the scintillation light by about 15 nanoseconds relative to the Cerenkov light. Also, because the positron typically travels only about 25 centimeters before it wears itself out, the scintillation light appears as an almost spherical cloud. The ratio of scintillator to mineral oil is selected to give roughly four scintillation photons to every Cerenkov photon. All told, a typical positron produces enough photons to trigger 450 phototubes.

Unlike the positron, which leaves a bright trail of light, the neutron goes quietly, wandering randomly away from the neutrino collision until it comes close enough to a proton to be captured through the reaction

$$n + p \to D + \gamma ,$$
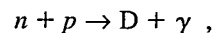
producing a deuteron (D) and a 2.2-MeV gamma ray (that is, a 2.2-MeV photon). On average, the capture takes 186 microseconds, so the 2.2-MeV gamma ray emerges somewhat after the light from the positron. The gamma ray also generates some scintillation light, which fires between 20 and 50 phototubes.

The light generates small electric pulses in the phototubes that then travel along some of the one thousand cables connecting the detector tank to the crates of electronics near the tunnel entrance. Like a brain processing a visual image, the electronics sift through the light signals, trying to assemble a picture of an electron antineutrino.

The purpose of the data collection electronics is to distinguish electron

Inside the LSND detector
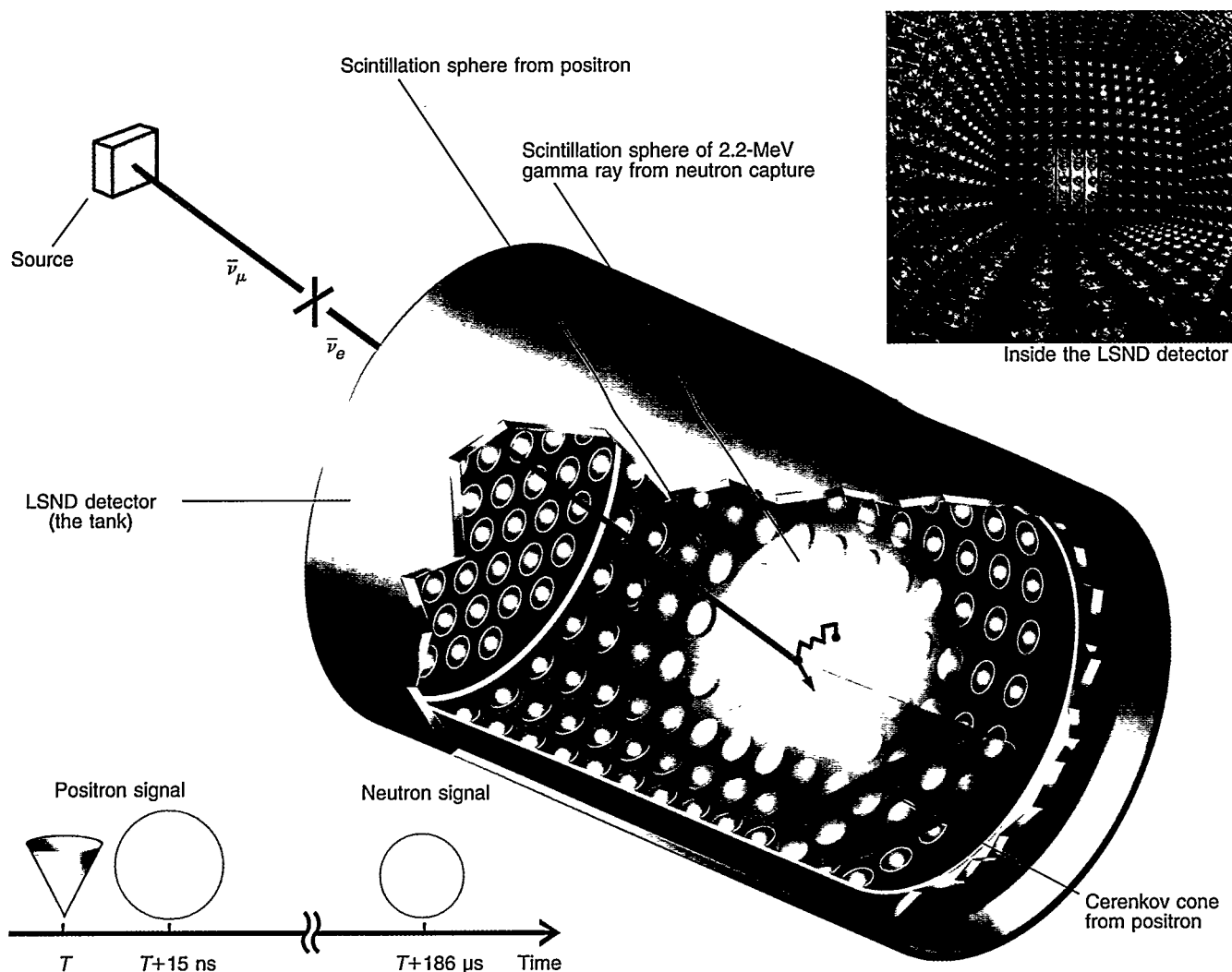
Figure 6. Signature of a Neutrino Event in LSND

In an oscillation, a muon antineutrino (blue) produced in the beam stop oscillates enroute to the detector and appears as an electron antineutrino (red). The neutrino strikes a free proton in the oil, creating a positron and a neutron. The positron travels faster than the speed of light in the oil and so produces a Cerenkov cone. As it loses energy through collisions with atoms in the oil, the positron also produces a sphere of scintillation light. The neutron survives about 186 microseconds and wanders 100 centimeters before it is absorbed by a nucleus, emitting a 2.2-MeV gamma ray that also produces a sphere of scintillation light. This succession of events—the apex of a Cerenkov cone centered on a sphere of scintillation light followed by emission of a 2.2-MeV gamma ray—is the signature of an electron antineutrino.

antineutrinos from the endless stream of cosmic rays that penetrate the detector's shielding and enter the tank. The probability that an electron anti-neutrino (or any other neutrino) will interact with matter is unimaginably small. Even with 167 metric tons of mineral oil, 99.999999999 percent of the neutrinos will pass through the tank unhindered and unnoticed. In a day, only 25 neutrinos will leave their mark.

In the same time period, 300 million cosmic rays will also pass through. (For more details on the data collection electronics, see "From Tank to Tape— The LSND Data Acquisition System" on page 112.)

The scenario is summarized in Figure 6. The electron antineutrino penetrates the tank and strikes a proton, giving rise to a neutron and a positron. The positron generates a cloud of blue

scintillation light and the characteristic cone from Cerenkov radiation. Then all is quiet for roughly 186 micro-seconds, after which a tiny 2.2-MeV gamma ray signals the presence of a neutron. Taken together, these signals constitute the signature of an electron antineutrino.

The central difficulty in analyzing the data is how to pick out the real electron antineutrinos. After all,

Figure 7. Identifying a Positron

An energetic particle moving through the tank creates both Cerenkov and scintillation light, but the ratio ρ of Cerenkov to scintillation light varies depending on the particle's type and energy. Positrons (or electrons) tend to have a very consistent ratio of about 0.5, which is evident in this figure as a large spike. Neutrons and protons tend to produce very little Cerenkov light, and thus they have small values of ρ (the rounded hump around 0.15). This pronounced difference in ρ is used to help distinguish positrons from other particles that leave trails in the tank.

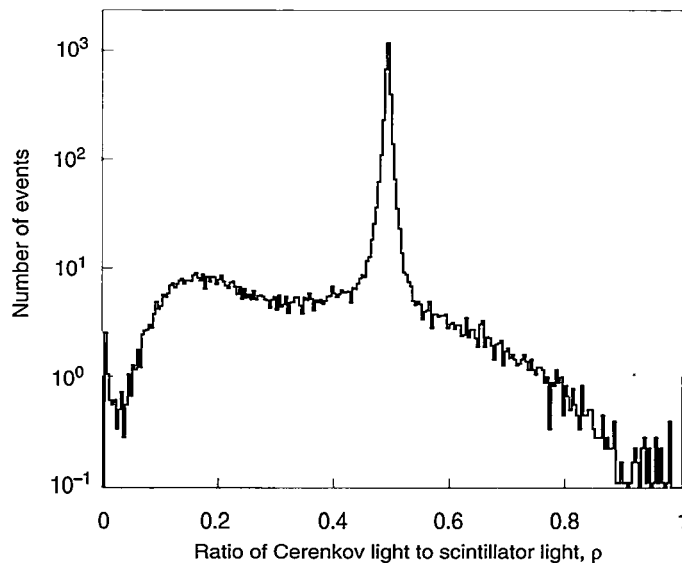Ratio of Cerenkov light to scintillator light, ρ

Figure 8. Accidental Photons

Photons due to background processes—accidental photons—should appear at random positions in the tank. However, these plots reveal that most accidental photons appeared at the bottom of the tank. Each point represents the position of a photon projected onto a two-dimensional plane. The heavy concentration along the bottom was likely due to a piece of steel shielding that has some minute amount of radioactivity or to a small hole in the shielding through which the electronic cables pass.

Horizontal position (cm)

Position along length of tank (cm)
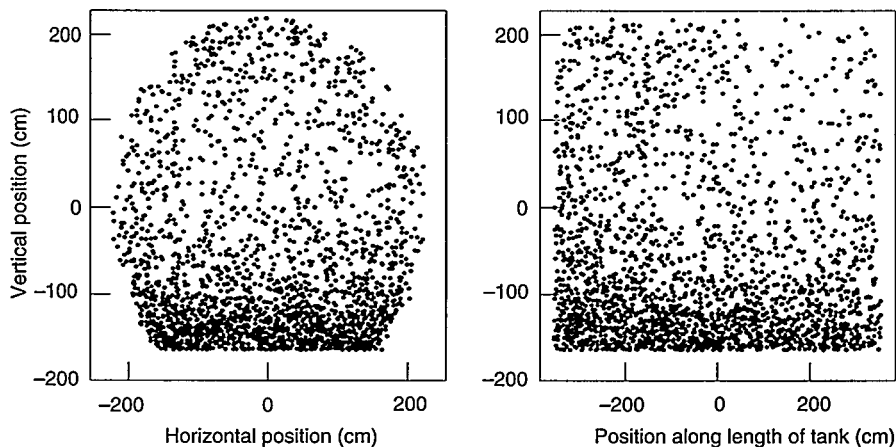
neutrinos aren't the only particles shaking up the atoms in the oil. Even with a carefully designed detector, events are not always what they seem. Pull the plug on the accelerator, and the detector will still record events that look indistinguishable from electron antineutrinos. As in any physics experiment, the researchers began by playing devil's advocate and drawing up an exhaustive list of possible impostors, or backgrounds. Topping the list were two potential showstoppers. First, what if the positron wasn't really a positron but was a cosmic ray instead? And second, what if the putative photon from neutron capture really came from somewhere else? To separate the real electron antineutrinos from the fakes, LSND researchers had to develop a variety of tools.

Positrons, Photons, and Impostors

The first problem was how to distinguish a positron from cosmic-ray muons and their by-products. Because most muons betray their identity by leaving a signal in the cosmic-ray veto shield (the outer shell that encloses the main detector tank), they are easy to track. If a muon is energetic enough to make it through the veto shield and into the tank, it tends to make a flashy entrance, producing so much ionization that all 1,220 phototubes light up. Looking more like a than a positron, these muons are easy to identify.

Occasionally, however, a cosmic ray can pull off a more convincing impersonation, passing near the tank and

knocking a neutron free in the shielding. The neutron can pass undetected through the veto shield and into the tank. There it may strike and propel a proton through the oil. If the proton is mistaken for a positron and a neutron is captured nearby, the combination would be a dead ringer for an electron antineutrino. (See "Other Things in the Tank: Backgrounds" on page 104.)

But forging a positron signature is not so easy. A typical 45-MeV positron lights up 450 phototubes, each of which measures the time and number of the photons it receives. This is an enormous amount of information from which the positron's position, trajectory, and energy can be culled. The positron's position can be determined to about 25 centimeters by finding the point in the tank that is equidistant in
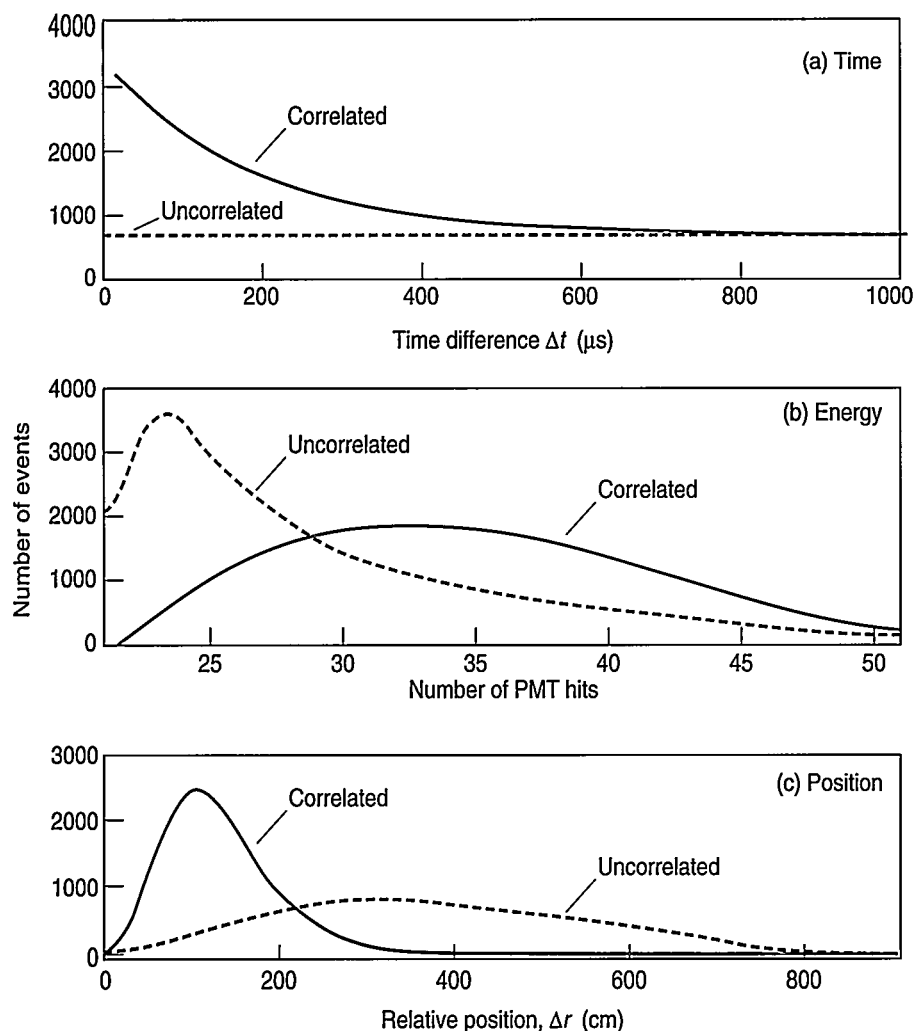
Figure 9. Identifying the Photon from Neutron Capture

Although the photon from neutron capture cannot be distinguished from a random photon, statistically, it has a distinct signature in time, energy, and position. In these graphs, blue lines show the distributions of photons that are correlated with neutrons, while red lines show the distributions for photons that are uncorrelated. (Cosmic-ray neutrons leave a telltale trace of light when they stop in the tank. They are used to determine the correlation functions.)
(a) On average, a neutron is captured 186 microseconds after it is created. Uncorrelated photons show no time correlation to the initiating event.
(b) Background photons have less energy and thus produce less light in the tank. On average, the 2.2-MeV photon lights up about 35 photomultiplier tubes (PMTs).
(c) The neutron wanders about 100 centimeters before it is captured. Photons arising from random processes can occur anywhere in the tank. These distributions are used in the likelihood function R to assess whether a photon is associated with a neutron created from a neutrino event.

time from all of the phototubes that detected photons. The trajectory can be determined by finding the ring of Cerenkov light that is superimposed on the uniform scintillation light. The trajectory runs through the center of the ring and can be determined to within 12 degrees. The positron's energy is simply proportional to the total charge from all of the phototubes that were hit and can be determined to within 6 percent.

Protons produced by cosmic rays, by contrast, will rarely be traveling fast enough to emit a Cerenkov cone (see Figure 7). Requiring a well-defined cone and an accompaning sphere of scintillation light removes 99.9 percent of all cosmic-ray events, while 80 percent of the real positrons pass the selection criteria.

The second problem was to sift out the photons (the 2.2-MeV gamma rays) that truly came from neutron capture. The steel of the cosmic-ray veto shield around the main detector tank absorbed most photons coming from the outside, but there was a chink in its armor. The shield covered the tank like an arch, stopping at the tunnel floor and leaving the underbelly of the detector exposed to the concrete. To extend the shielding, the LSND collaborators had laid down a 15-centimeter-thick floor of steel planks underneath the detector.

Even with this flooring, however, accidental photons turned out to be a bigger problem than expected. When the LSND physicists looked for photons in their 1993 data, they found many of them clustered suspiciously at the bottom of the tank toward one end

(see Figure 8). Whether due to a bit of slightly radioactive steel or to the small hole in the shield where the electronic cables exited the tank, the extra photons threatened to swamp the real electron antineutrino signature.

To filter out these accidental photons, Richard Imlay and his group from Louisiana State University came forward with a new technique based on a quantity called R, which determined the likelihood that the "positron" and "neutron" signals were correlated. If the two were the true signature of an electron antineutrino, the photon (from neutron capture) should have 2.2 MeV of energy and appear slightly after the positron but at roughly the same location in the detector tank (see Figure 9).

This correlation was in contrast to what you expected for an unrelated or

Other Things in the Tank: Backgrounds

Despite the shielding that surrounds it, the detector tank remains a bubbling cauldron of activity. Every second, thousands of accidental photons and cosmic rays stream through the oil, leaving trails of scintillation light in their wake. Occasionally, the endless background signals combine to look just like the signature of an oscillation event—a positron followed some 186 microseconds later by a photon. The LSND experimenters have taken great pains to understand how background signals could mimic real electron antineutrino events so they can estimate the number of "false positives" that would make it through the data analysis. Only by finding an excess of oscillation over background events can they state with confidence that they have observed neutrino oscillations.
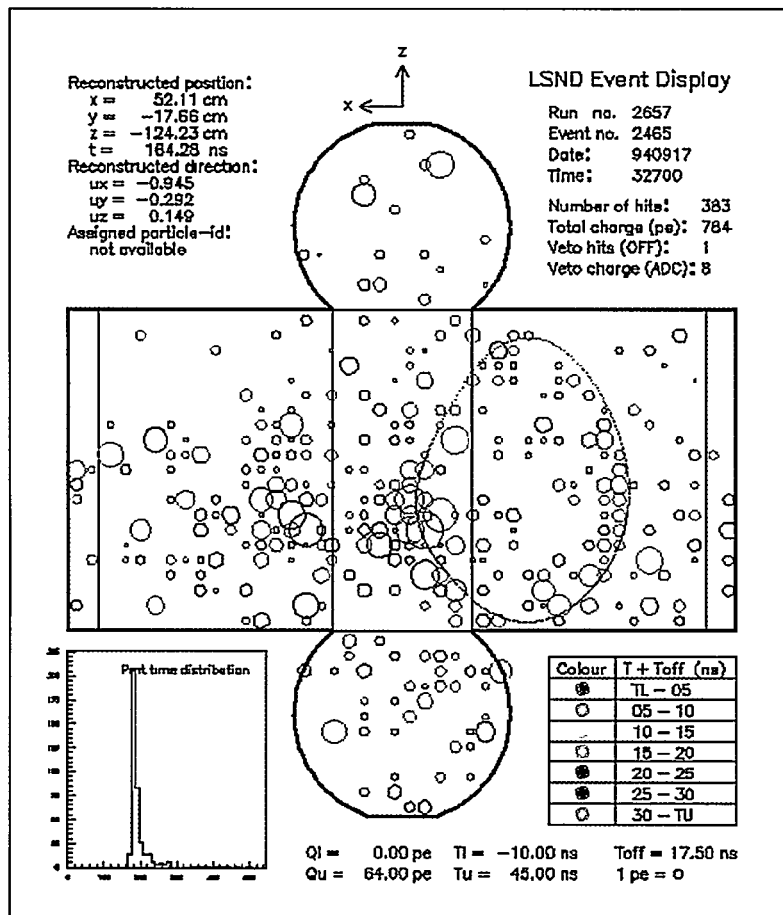
Non-Beam-Related Backgrounds: Cosmic Rays and Accidental Photons. Cosmic rays are the largest source of background. If every muon antineutrino were to oscillate to an electron antineutrino, cosmic rays interacting in the tank would still outnumber electron antineutrinos by a factor of more than 100,000. Most cosmic rays that reach the earth are muons, produced from the decay of pions created when high-energy protons strike nuclei in the upper atmosphere.

Surrounding most of the detector tank is a cosmic-ray veto shield that is LSND's main line of defense. The archlike shield has double walls. The inner wall is a 15-centimeter-thick layer of lead shot that absorbs accidental photons and a significant number of cosmic rays. But this dense layer is not enough. Every second, approximately 4,000 cosmic-ray muons pass straight through the lead and enter the tank. Thus, the outer wall of the veto shield is studded with 292 photomultiplier tubes looking inward, and the space between the walls is filled with mineral oil and liquid scintillator. On their way into the tank, cosmic rays leave a trail of scintillation light in the veto shield. Removing events in which more than a few of the veto-shield phototubes fire eliminates 99.999 percent of all cosmic-ray-induced events.

Reconstructing an LSND Event. In this flattened view of the tank's inner surfaces, the colored circles identify 383 photomultiplier tubes that have been hit by photons. A circle's diameter corresponds to the number of photons that struck the tube; its color indicates when those strikes occurred (see color key at the lower right). This data is used to reconstruct the Cerenkov ring (which looks oval-shaped in this projection) and the scintillation sphere (not shown). The position and trajectory of the particle associated with the photons (see upper left) are then calculated. In this case, the particle was an energetic electron produced from the decay of a muon.

Being very energetic, most cosmic-ray muons that pass through the shield also pass straight through the tank. About 10 percent, however, don't make it. They stop in the tank and decay, with the positive muons producing positrons and the negative muons producing electrons. (The two particles are treated the same in the data analysis, since they are indistinguishable as seen through the eyes of the detector.) Although the muon is detected by the veto shield, the positron that is born typically 2 microseconds later is not. It appears to come from nowhere, exactly like the positron created from a electron antineutrino event. Thus, in this background process, the muon acts as a kind of Trojan horse for the positron, in effect sneaking it past the defenses of the veto shield.

Most of these positrons, however, are ignored by the data acquisition system (see the box "From Tank to Tape—The LSND Data Acquisition System" on page 112). The system requires that there be no activity in the detector or veto shield for a period equivalent to

LSND Event Display

Reconstructed position:
x = 52.11 cm
y = -17.66 cm
z = -124.23 cm
t = 184.28 ns

Reconstructed direction:
ux = -0.945
uy = -0.292
uz = 0.149

Assigned particle-id:
not available

Run no. 2657
Event no. 2465
Date: 940917
Time: 32700

Number of hits: 383
Total charge (pe): 784
Veto hits (OFF): 1
Veto charge (ADC): 8

Print time distribution

Colour	T + Toff (ns)
	TL – 05
	05 – 10
	10 – 15
	15 – 20
	20 – 25
	25 – 30
	30 – TU

Ql = 0.00 pe   Tl = -10.00 ns   Toff = 17.50 ns
Qu = 64.00 pe   Tu = 45.00 ns   1 pe = o

about 7 muon lifetimes before a positron appears. If any activity occurs during this "all quiet" time, the positron is rejected as signaling a potential neutrino event.

Although not all cosmic-ray background events can be detected, their number is relatively easy to estimate: it is simply measured when the accelerator is off. The accelerator does not produce protons continuously but has a regular heartbeat, pumping out 600-microsecond bursts of protons 120 times a second. The beam is on only about 7 percent of the time, and the non-beam-related backgrounds can be studied during the relatively long rest periods that compose the remaining 93 percent. In fact, the data acquisition system does not initially distinguish between beam-on and beam-off events. It only uses that information later, when it assesses which events represent true neutrino interactions. In their 1996 paper, the LSND collaborators estimated that they expected 2.5 ± 0.4 events from beam-off sources to look like electron antineutrinos (Athanassopoulos et al. 1996).

Beam-Related Backgrounds: Neutrinos. The beam contains equal numbers of muon neutrinos, muon antineutrinos, and electron neutrinos. If, for instance, 1 percent of the muon antineutrinos oscillate to electron antineutrinos, they will be outnumbered 300 to 1 by the other neutrino types. Although the other neutrinos cannot easily imitate electron antineutrinos, they can still lead to background events that must be estimated.

A muon antineutrino interacting with a proton produces a positive muon (instead of a positive electron) and a neutron. This process is a potentially dangerous background; the muon can decay in the tank and produce a positron, which, when combined with the neutron, would convincingly mimic the oscillation signature of an electron antineutrino. Fortunately, the newly created muon produces Cerenkov and scintillation light in the tank, so the muon is observed by the data acquisition system. The "all quiet" requirement removes these muon decay events. In addition, because the muon weighs a hefty 105 MeV, only the relatively few decay-in-flight muon antineutrinos have enough energy to produce muons. (All decay-at-rest neutrinos have energies below 55 MeV.) In the end, muon antineutrinos constitute a small background that can be reliably calculated.

Electron neutrinos are a background because they can change a carbon atom in the detector's mineral oil into a nitrogen atom. An electron is also produced in the process:

$$\nu_e + C \rightarrow e^- + N \ .$$
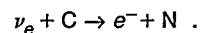
However, since no neutron is produced in this reaction, the process is only a problem in the unlikely event that it coincides with an accidental photon.

The largest source of beam-related backgrounds is the electron antineutrinos present in the beam itself. They arise from negative muons that decay in the beam stop before being absorbed. These decay-product neutrinos are essentially indistinguishable from electron antineutrinos produced from oscillations. Fortunately, the decay-product flux is very well known, and this background can also be calculated with confidence. In their 1996 paper, the LSND collaborators estimated the background from electron antineutrinos in the beam to be 1.1 ± 0.2 events and the total background from beam-related events to be 2.1 ± 0.4 events. Thus, all told, backgrounds accounted for a grand total of 4.6 ± 0.6 events (Athanassopoulos et al. 1996).

"accidental" photon, which could come from the small amounts of radioactive elements (thorium, for instance) present in the concrete and earth surrounding the detector or in components used to build the phototubes. Accidental photons tended to have energies below 2.2 MeV and, by definition, appeared uncorrelated in position and time with the positron. The data plots shown in Figure 9 reveal some of the differences between accidental photons and those associated with neutron capture.

The R correlation not only picked out events in which the positron and photon were correlated, but it also rejected events in which the photon looked accidental. Mathematically speaking, R was a "relative likelihood," but it worked like a magic box. Feed in the number of phototubes fired by the photon and the distance and time between the photon and the positron signals, and out popped a number. Real electron antineutrino events tended to have high values of R, while accidental photon events piled up at low R. Unfortunately, although R could remove 99.4 percent of the accidentals, it did so at a high price: R also removed 77 percent of the real electron antineutrinos.

The final source of background events was neutrinos produced in the beam stop. The electron antineutrinos that contaminated the source contributed a few background events, as did the other neutrino types that could, on occasion, leave what appeared to be the signature of an electron antineutrino.

Still, when the LSND collaborators ran through the data from both 1993 and 1994, they found further evidence for electron antineutrinos. Their improved analysis let fewer impostors slip through the cracks than before. This time they had nine events that looked like electron antineutrinos. The expected number of background events came to only 2.1 ± 0.3. The odds, they calculated, that background could account for the nine events were roughly 1 in 300.

One Experiment, Two Interpretations

These were heady times at LSND, but they were also times of trouble: R had driven a wedge between some members of the group. Alfred Mann, one of the original collaborators and a professor at the University of Pennsylvania, withdrew from the collaboration over what he saw as a dangerous disregard for the scientific method. Mann feared that the group had lost objectivity. The experimenters wanted to see an excess of events, he believed, so they had unconsciously shaped their analysis to find one. Mann was an authority, having worked on many experiments designed to search for the as-yet unobserved, and he preferred simpler methods to the "unnecessarily complex" R. "My whole experience says that if you're going to find something new, it generally rises up out of the data and pokes you in the eye," he said.

At the same time, one of Mann's graduate students, James Hill, had returned to Philadelphia from Los Alamos and was hard at work on his own analysis of the LSND data. Hill decided to restrict himself to the 1994 data and to dispense with R. It made more sense, he argued, to simply cut out photons that appeared in the bottom of the tank or near the tank walls. These cuts, however, reduced the data set by a factor of 3. Hill also used less-restrictive selection criteria to pick out photons. In the end, he found five events that looked like electron antineutrinos. He calculated the background to be 6.2 ± 1.6 events. By his estimate, the excess was a mirage.

But the two analyses weren't necessarily contradictory. If there were only a few oscillation events, it was entirely possible that they would stand out in one analysis but not in another—especially, the advocates of R argued, if the other analysis cut out two-thirds of the data and took a simple-minded approach to selecting photons. Mann, on the other hand, thought Hill's work was "entirely sensible." It was,

he said, the conservative and hence safe approach. The debate was amicable, Mann maintains: "We just differed in our interpretation of the data."

When January 1995 rolled around, LSND again found itself on the losing side of a budget war. DOE had begun to draft a five-year plan that, in the words of one collaborator, "slit the throat of LSND." The collaborators decided it was time to show their hand. They would go to a nuclear and electroweak physics meeting at Berkeley later that month and announce that they had what looked like a hint of oscillations. Since it would be impolite, they reasoned, not to present the results first at Los Alamos, they scheduled an on-site colloquium in advance. Bill Louis, the LSND spokesman, would give the Los Alamos colloquium on Thursday. Hywel White would give the Berkeley talk the following Sunday.

Fit to Print

With preparations underway for the Thursday colloquium, the phone rang with a call that would change everything. It was the *New York Times*. Someone, perhaps at a recent astrophysics conference, had tipped *Times* reporter John Wilford to the LSND results. Wilford called John Gustafson at the Los Alamos public affairs office, and Gustafson, pleased with the prospect of a *New York Times* article, approached the LSND group with the idea of letting the *Times* report on the results. The collaborators on hand at Los Alamos were hesitant about speaking to the press before informing their peers, but in the end they decided that it was better to talk to Wilford than to let the *Times* run a story that could be wrong or overblown. White recalls a sense of helplessness: "It's just like going on a rubber raft," he remembers. "Once you decide to get on, jumping off doesn't make any sense, so you hang on as best you can."

The ride was long and rough, and in retrospect, many wished they had kept

to higher, drier ground. The Tuesday before the Los Alamos colloquium, the *Times* ran its story—not buried behind the fashion page in the Science Times section but on page one, just below the fold, making it look like a definitive discovery. "Cosmos's Missing Mass: Wispy Particle Weighs In," the headline read (Wilford 1995). Looking at the article now, White says it seems balanced and accurate, but at the time, it made him swallow hard. In the physics community, there is nothing as close to a sin as "publishing in the *Times*." Colleagues want to have a crack at reviewing new results before they hit the press. "They really screwed the pooch," one physicist remembers thinking after reading the article at breakfast. The incident gave critics a peg to hang their skepticism on.

In the following weeks, LSND collaborators found themselves apologizing for their misstep with the media. "If I could do it again," says Louis, "I would just say 'no comment.'"

At the same time, word of Hill's contrary analysis began to circulate, fueling doubt in the physics community as to the credibility of the results. Louis recalls, "They seemed to be saying 'we don't know what you've got there, but it's not oscillations.'" One physicist at a Fermilab colloquium said White seemed slick, "like a lawyer who knew his client was guilty." Vern Sandberg, an LSND experimenter, says that under the circumstances, he could sympathize with the skeptics: "I wouldn't have believed us either."

Looking back, Imlay thinks the results were a hard sell because few in the audience had the expertise to understand them. And it was true, LSND sat at the intersection of two, vast fields—high-energy physics and nuclear physics—fields that were like adjacent neighborhoods that spoke different languages. "Neutrino physics is a niche," Imlay says, "it's not the sort of thing where you can walk in, hear a talk, and understand it." Things would have gone easier, he suspects, if the group had first published some conven-
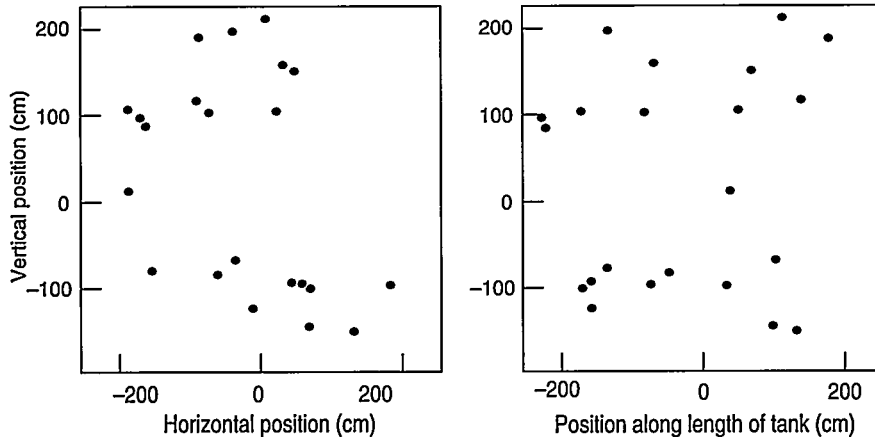
Figure 10. Evidence for Oscillations
The likelihood function R was used to sift through the 1995 LSND data to determine if a photon was likely to have come from neutron capture and was correlated with a positron event. In all, there were 22 events when the beam was on that had R values >30 (that is, that had a high level of correlation). Only 4.6 ± 0.6 events were expected from background processes. The 22 events appeared scattered across both tank cross sections.

tional results to establish credibility. But there had been no time for such niceties.

Doubt even began to trouble the collaborators as they prepared their paper for publication. The process, which by some estimates should have taken two weeks, took two months. Right up until the final edit, Louis recalls getting a flood of comments by E-mail. Had we checked this? Could we change this word here? Hill also wanted to publish his results, and by this time, the relationship between Hill, Mann, and the rest of the LSND group had become so strained that the two factions were unable to consolidate their results into a single paper. In April, they submitted two papers to *Physical Review Letters*, which ultimately ran them back to back. One bore the names of 39 collaborators; the other had a single author, James Hill (C. Athanassopoulos et al. 1995, Hill 1995).

In politics, debates can drag on indefinitely, without hope of resolution. In physics, there is no arguing with nature. With more data, the truth would out. Eventually.

Back Out to Sea

Fortunately, it looked like there would be more data. Though the nuclear physics division of DOE had orphaned LANSCE (which at the time was called LAMPF for Los Alamos Meson Physics Facility) by cutting its

funding in 1995, the defense projects division had arranged to assume custody. The new overseers renamed the accelerator LANSCE to reflect what would be its new focus on neutron physics but agreed to allow neutrino production to continue on the side.

Gerry Garvey, an LSND collaborator who was watching events from a temporary post at the White House's Office of Science and Technology, attributed LANSCE's stay of execution to a law of nature. "Good things have a way of continuing," says Garvey, who had been LAMPF's director for five years; "names may change, people will come and go, but where there is will, research will persevere."

Drawing upon outside sources and Los Alamos discretionary funds, LSND cobbled together enough money to run for another four months. The much-anticipated run began in August 1995. That first month, Louis checked every day to see if the detector had recorded any electron antineutrino events. They expected less than one per week. It was like being on a long fishing expedition, staring into the dark waters and waiting for something to bite. But every day the nets came up empty.

Louis began to fear that they might have been wrong. Anxiety woke White at 2 A.M. many mornings to think the experiment through again. Insomnia caught on like the flu, leaving many weary and frustrated. Some questioned the electronics that stood between the

collaborators and the neutrinos. Questions popped up like "We've got 1,500 channels here, could there be some mistake? A dirty connection? A glitch in the trigger memory? An electronic hallucination?" But Sandberg, the technological hero who had engineered the data acquisition system, had a parental faith in its performance. Every day he dove into the raw data and made sure the hardware was doing what it should. Still, Sandberg remembers having his doubts, too: "We were worried we might end up in the Journal of Irreproducible Results."

Then on the last day of August, a single event came in. In September there were a few more. At the end of the run, looking at the entire data set, they had a grand total of 22 events with a predicted background of 4.6 ± 0.6 (see Figure 10). "I learned a new appreciation for what low statistics means," recalls Imlay, salvaging a lesson from the nail-biting experience. The group estimated that the odds that all 22 events were background were less than 1 in 10 million. Working backwards, the collaborators calculated the possible regions of Δm^2 and $\sin^2 2\theta$ that could explain the oscillations. Many had been ruled out by previous experiments, but a few small regions stretched tantalizingly out into the unprobed region.

Burned by the spotlight once, the collaborators stonewalled their curious colleagues and took a full five months
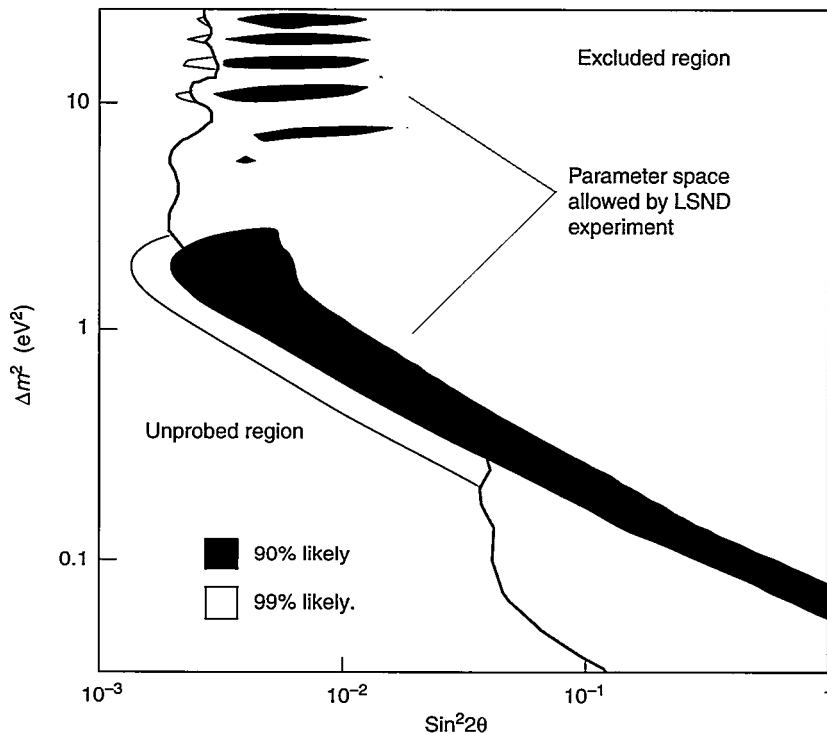
Figure 11. Parameter Values from the Initial Analysis of the LSND Data

Much of the area favored by the LSND search (blue regions) had been ruled out by previous searches (the grey region). A small strip and the edges of a few of the "islands" appeared to be the most likely values of Δm^2 and $\sin^2 2\theta$ that could have generated the oscillations. The true parameter values are 90 percent likely to lie in the dark blue region and 99 percent likely to lie in the light blue region. The shape of the allowed regions is caused by the fact that two terms in the oscillation probability expression $P(\bar{\nu}_\mu \rightarrow \bar{\nu}_e)$ can vary independently.

preparing their next paper. "People would say, 'Hello! How are you? Haven't seen you in a while, so . . . WHAT ARE THE RESULTS?'" Sandberg remembers, "but we made them wait." The idea was to write a paper that would hang together with the certainty and logic of a mathematical proof. Publicity had forced intense introspection, and this time the group produced a 24-page tome entitled "Evidence for Neutrino Oscillations from Muon Decay at Rest" (C. Athanassopoulos et al. 1996). A colleague praised it as one of the most extensive and exhaustive descriptions of an analysis ever published. Around LSND, it is simply referred to as "the big paper."

The big paper spelled out the analysis in excruciating detail, at the level of

a graduate thesis, and went a long way toward restoring LSND's credibility. Paranoia had paid off.

In the paper, the collaborators also tried cutting out photons in the bottom of the tank as Hill had done. They found six events with a background of 1.7 ± 0.3 events. The odds were about 1 in 100 that the six events could all be background. Responding to Mann's allegations that they had willed the excess into existence, they repeated their analysis, varying the requirements for photons and positrons. In each case, they reported an excess of events.

But the physics community reasonably demands a high level of proof before declaring victory. Today Mann, while impressed by the excess, cautions that the results are nothing to yell Eureka about: "If you knew your house had a

1 percent probability of burning down, you'd be out buying insurance," he says.

Mapping out the Territory

Nothing makes a physicist happier than data, and having observed a statistically significant excess of events, the LSND collaborators proceeded to map out the regions of Δm^2 and $\sin^2 2\theta$ that could have led to the oscillations.

Given the detection efficiency, they could calculate the number of electron antineutrinos that had passed through the tank without getting caught. By comparing this number with the number of muon antineutrinos that emanated from the source, they estimated that only 0.31 ± 0.13 percent of the muon antineutrinos had oscillated. To unfold the Δm^2 and $\sin^2 2\theta$ information from the data, the collaborators performed a likelihood fit to all events that contained a positron. The fit took into account the positron's energy and direction, the photon likelihood R, and the distance to the source for each event. It also took into account the expected distributions of these quantities for electron antineutrinos from oscillations and for background processes.

As shown in Figure 11, the fit does not pinpoint particular values of Δm^2 and $\sin^2 2\theta$ but rather carves out regions of more-likely values. The shapes of the regions are a consequence of the fact that the oscillation probability is the product of two terms, one relating to Δm^2, x, and E, and the other to $\sin^2 2\theta$. The spots that spread out like a chain of small islands arise because some oscillation events at relatively high energy tend to exclude Δm^2 near integral multiples of 4.3 eV^2. (Those values of Δm^2 make multiples of the oscillation wavelength approximately equal to the source-to-detector distance. Hence, $\sin^2(\pi x/\lambda)$ is almost zero.) The longer "island" corresponds to smaller values of Δm^2, where $\sin^2(\pi x/\lambda)$ no longer oscillates but slowly approaches zero and where the data must be accounted for entirely by increasing the allowed values of $\sin^2 2\theta$.

Figure 12. LANSCE Production of Decay-in-Flight (DIF) Neutrinos and Their Energy Spectra

$$\pi^+ \rightarrow \mu^+ + \nu_\mu$$

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$$

(a)
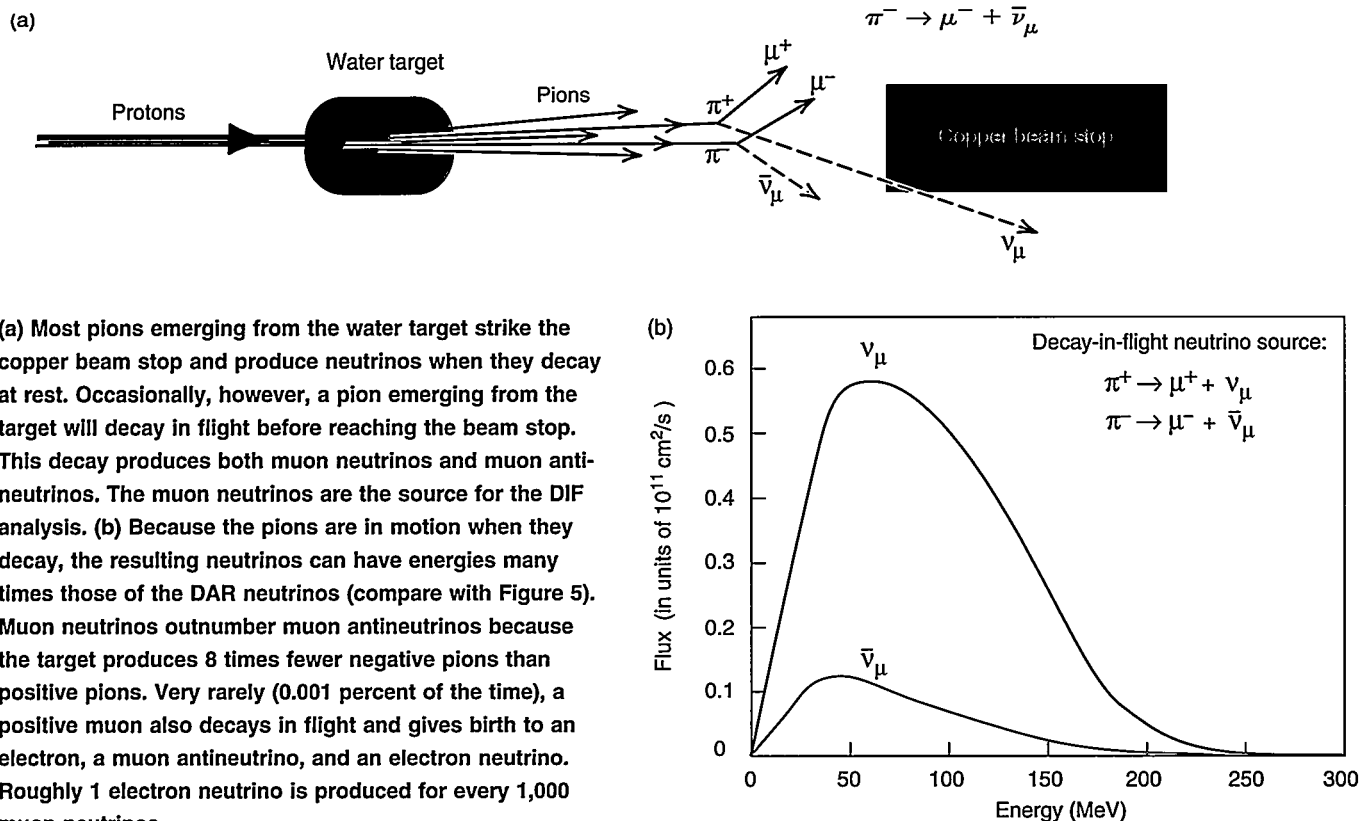
(a) Most pions emerging from the water target strike the copper beam stop and produce neutrinos when they decay at rest. Occasionally, however, a pion emerging from the target will decay in flight before reaching the beam stop. This decay produces both muon neutrinos and muon anti-neutrinos. The muon neutrinos are the source for the DIF analysis. (b) Because the pions are in motion when they decay, the resulting neutrinos can have energies many times those of the DAR neutrinos (compare with Figure 5). Muon neutrinos outnumber muon antineutrinos because the target produces 8 times fewer negative pions than positive pions. Very rarely (0.001 percent of the time), a positive muon also decays in flight and gives birth to an electron, a muon antineutrino, and an electron neutrino. Roughly 1 electron neutrino is produced for every 1,000 muon neutrinos.

(b)

Many of the regions allowed by the LSND data, however, had already been explored and found barren (see Figure 4). An experiment at Brookhaven National Laboratory (E776) and a French reactor-based experiment (BUGEY III) had essentially ruled out all of LSND's preferred areas except for a narrow strip that stretched from $\Delta m^2 = 0.2$ to 20 eV^2 and from $\sin^2 2\theta = 0.03$ to around 0.001. Although the LSND results taken alone allowed a variety of interpretations, when combined with the previous null results, the parameter space for neutrino masses and mixing angles was quite limited.

Decay in Flight—The Second Analysis

While the LSND collaborators were drawing up Figure 11 and putting the finishing touches on the big paper, they

began work on a second analysis that would either confirm or disprove all their previous work. Knowing that an excess of electron antineutrinos was not enough to establish that oscillations had occurred (the electron antineutrinos could be coming from some equally surprising source, such as an exotic type of muon decay), the collaborators had built in a second method to look for oscillations. Instead of using the neutrinos from decay-at-rest (DAR) pions and muons, this method looked to a smaller sample of neutrinos produced by the pions that decayed in air on their way to the beam stop. These decay-in-flight (DIF) neutrinos had a much higher energy than the DAR neutrinos and so could be easily distinguished in the detector.

In the production of DAR neutrinos, pions produced in the target travel about a half meter through open space before striking and coming to rest in

the copper beam stop (refer to Figure 5). When the still pion decays, it can give at most 53 MeV, or about half of its mass, to one of the resulting neutrinos. Thus all neutrinos used in the DAR analysis had energies below 53 MeV. By contrast, a pion that decays in flight before reaching the beam stop passes on some of its kinetic energy to the resulting neutrino, giving it as much as 300 MeV of energy. By considering only neutrinos that had at least 60 MeV of energy, LSND collaborators could essentially tap a second neutrino source, for free. Figure 12 shows this DIF source and the energy spectrum of the neutrinos produced by it.

Unlike the DAR neutrinos, DIF neutrinos come mainly from pion decay, since there is rarely time in the half-meter journey to the beam stop for the muon produced by the pion to also decay. (Roughly 3 percent of the pions
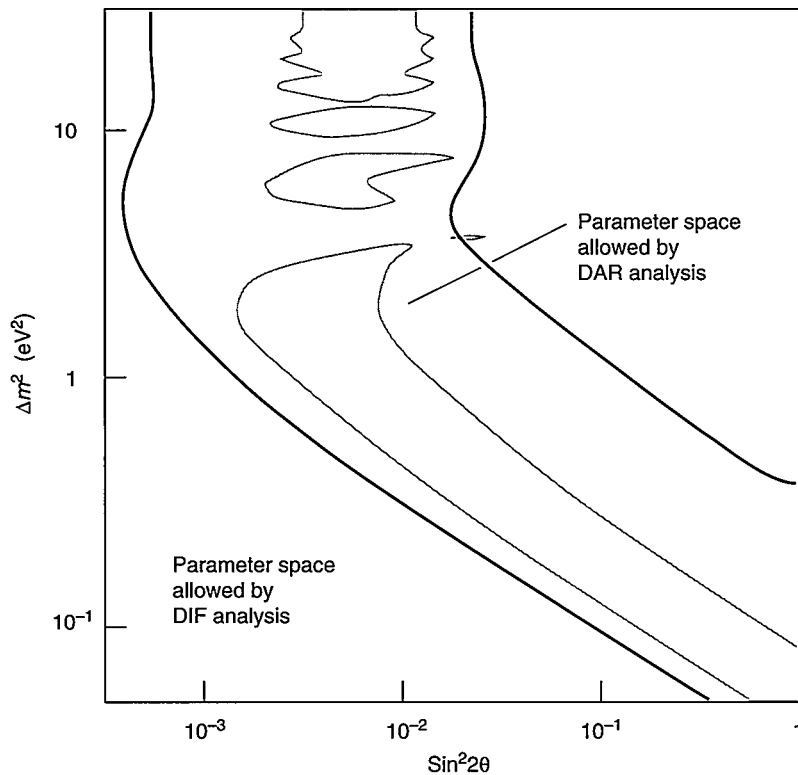
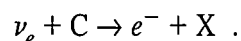## Figure 13. Parameter Values Allowed by DAR and DIF Neutrino Analyses

The DIF analysis also indicated neutrino oscillations: a small excess of electron neutrino events was detected above background events. The parameter values allowed by the DIF analysis lie between the two solid lines, whereas the values allowed by the DAR analysis are shown in grey. Although there were greater uncertainties associated with the DIF analysis, the allowed values of Δm^2 and $\sin^2 2\theta$ for the two analyses overlapped.

down to the difficult task of separating electrons from background sources such as cosmic rays.

Since the detector is essentially charge-blind, identifying electrons is just like picking out positrons in the DAR analysis. But unlike the positrons, the DIF electrons can have considerable energy, so much so that they may travel for half a meter before stopping in the oil. As a result, the "sphere" of scintillation light the electron produces in the tank becomes stretched out, looking instead like the superposition of spheres from a string of electrons.

The LSND group developed two methods for selecting electrons. Both made careful study of the amount and timing of the light expected to hit each phototube. Both also looked for a Cerenkov cone and scintillation light and discriminated against cosmic rays, although in slightly different ways. Finally, both methods offered improved position and direction resolution over the positron method used in the DAR analysis. In the end, each method had its own strengths and weaknesses, so the collaborators decided to use both. Taken together, the two methods were expected to pick out roughly 17 percent of the electrons produced by electron neutrinos.

The collaborators identified 40 events in the data set that seemed to contain a high-energy electron. They had expected roughly 11 events from backgrounds such as cosmic rays and 10 events from electron neutrinos present in the beam, leaving an unexplained 19.2 ± 7.8 events. They calculated the probability that backgrounds could account for the excess to be less than 1 percent. More importantly, when they worked backwards to see what neutrino masses and mixing angles could have generated the oscillations, the results overlapped quite nicely with the earlier DAR results (see Figure 13). The two analyses seemed to point to the same conclusion—that oscillations and massive neutrinos were behind the excess events.

decay in flight, but only 1 in 100,000 muons do.) As a result, the DIF neutrinos are mostly muon neutrinos and contain relatively few contaminating electron neutrinos. The DIF analysis looked for muon neutrinos to oscillate into electron neutrinos—the "matter" counterpart of the DAR analysis. If the DAR analysis observed oscillations, so too should the DIF analysis.

Alfred Mann thought the DIF analysis so important, and so integral to the mission of LSND, that he had urged the collaborators to keep a low profile until they finished it. But the LSND group had gone out on a limb with their DAR analysis, and it remained to be seen whether the DIF data would support the earlier analysis.

While the DIF analysis sought to observe the same kind of oscillations as the DAR analysis, it did so in a way that was completely independent. Even the way the detector observed the oscillations was different. The electron antineutrinos in the DAR analysis struck a free proton in the oil and produced a low-energy positron and a neutron. By contrast, the electron neutrinos in the DIF analysis interacted with a neutron in a carbon atom from the oil, producing a high-energy electron and transforming the carbon into another atom (X), typically nitrogen:

$$\nu_e + C \rightarrow e^- + X \ .$$

Unlike in the DAR reaction, no neutron is produced, so the DIF analysis boiled

Epilogue

At the time of this writing, LSND had just begun to make its DIF results public. This time around, reactions are more enthusiastic than condemning, and question-and-answer sessions no longer run an hour. Whereas before, LSND results were often downplayed in neutrino talks, they now take center stage along with the atmospheric and solar data.

Neutrinos defied detection for nearly 25 years after Pauli first proposed them, and today, near the 70-year mark, physicists still disagree over whether neutrinos have mass. While there seems to be a growing suspicion in the physics community that neutrinos do indeed have mass, many are still waiting for the day when some experiment sees the cyclic rise and fall of the number of neutrinos from oscillations as the neutrino energy or the distance between the source and the detector is gradually altered. It would be impossible to stare those results in the face and deny that neutrinos have mass.

That day would mark the end of one of the longest quests in the history of particle physics, one that currently stretches over half a century and spans generations of physicists. It would also reserve a place in the history books for LSND, solar, and atmospheric experiments. But even those intimately involved in neutrino work are uncertain exactly when all this might come to pass. White has a page from a word-a-day calendar tacked to the wall of his office, just beside a copy of the troublesome *New York Times* article. The word is obscure: *Greek calends*, defined as a time that will never arrive, the next blue moon, or when pigs fly. With the mounting evidence, it may be that pigs are preparing for takeoff. ∎

Further Reading

Athanassopoulos, C., et al. 1995. Candidate Events in a Search for $\bar{\nu}_\mu \to \bar{\nu}_e$ Oscillations. *Physical Review Letters* 75: 2650.

Athanassopoulos, C., et al. 1996. Evidence for Neutrino Oscillations from Muon Decay at Rest. *Physical Review C* 54: 2685.

Athanassopoulos, C., et al. 1997. Evidence for $\nu_e \to \nu_\mu$ Oscillations from Pion Decay in Flight. Submitted to *Physical Review C*. LA-UR-97-1998.

Hill, J. E. 1995. An Alternative Analysis of the LSND Neutrino Oscillation Search Data on $\bar{\nu}_\mu \to \bar{\nu}_e$. *Physical Review Letters* 75: 2654

Louis, W. C., et al. 1994. Neutrino Oscillation Studies at LAMPF. *Nuclear Physics B* (Proc. Suppl.) 38: 229.

Mukerjee, M. 1994. Missing Matter Found? *Scientific American* 271 (2): 22.

Wilford, J. 1995. Cosmos's Missing Mass: Wispy Particle Weighs In. *The New York Times*. January 31, 1995.

William C. Louis is currently the scientific spokesman of the liquid scintillator neutrino detector (LSND) experiment at LANSCE. Following a postdoctoral appointment at Rutherford Laboratory and a faculty appointment at Princeton University, Louis joined the Laboratory in 1987 as a staff member of the Medium Energy Physics Division. In 1993 he joined the Physics Division's Subatomic Physics Group. Louis earned his Ph.D. in physics from the University of Michigan in 1978, studying high-energy neutrino interactions at the then recently opened Fermi National Accelerator Laboratory. Louis' research at Los Alamos has centered on weak interactions, neutrino reactions in nuclear and particle physics, and neutrino oscillations.

Vern D. Sandberg received his Ph.D. in 1975 from the University of Utah, where he studied strong gravitational fields. Following a postdoctoral appointment at the Center for Relativity Theory at the University of Texas at Austin, Sandberg accepted a postdoctoral position at CalTech, where he participated in the discovery of quantum non-demolition techniques that circumvented the "quantum limit" for the readout of gravitational wave detectors. In 1979, he joined LAMPF (now LANSCE), working with Darragh Nagles to develop instruments for precision experiments. The high intensity of LAMPF's accelerator beam opened up possibilities of experiments with well-defined beams of neutrinos, and soon thereafter he collaborated on the neutrino-electron elastic scattering experiment that measured the interference between the weak neutral current and the charged neutral current in the scattering of electrons by neutrinos. This experiment led to a program to search for neutrino oscillations, which culminated with the LSND experiment. Sandberg is currently involved in the neutrino oscillation physics program, consulting on gravitational wave detectors and building a quantum computer.

D. Hywel White received his B.Sc. in mathematics and physics from the University of Wales in 1953, and earned his Ph.D. in experimental particle physics from Birmingham University, England, in 1956. After a stint on the faculty at Birmingham University, White became an assistant professor at the University of Pennsylvania. In 1964 he was appointed associate professor, and in 1968 professor, at Cornell University. White joined Brookhaven National Laboratory in 1978 as head of experimental facilities on the ISABELLE project, where he began experiments in neutrino physics. In 1986 he came to the Laboratory as group leader of Nuclear and Particle Physics Research at LAMPF to continue work in experimental neutrino physics. In 1994 the neutrino group observed a signal interpreted as neutrino oscillations, an event that has since dominated their research.
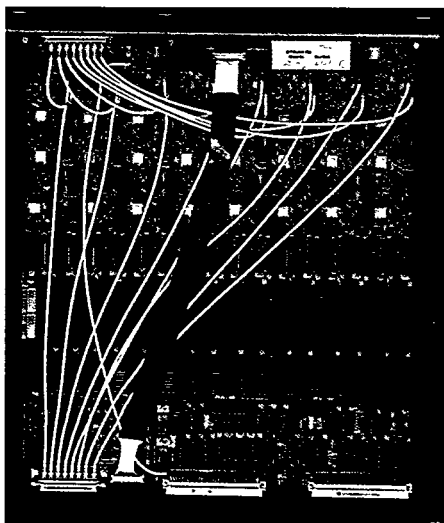
From Tank to Tape—The LSND Data Acquisition System

Neutrinos interact in the LSND tank at a rate of approximately one an hour. Go to lunch, and if you're lucky, there will be a new neutrino event written to tape when you get back. But in that same hour, nearly 15 million cosmic rays will also have left their marks in the tank. That's a staggering number to contemplate. "You have all this background from cosmic rays," says Vern Sandberg, the principal designer of the LSND data acquisition system, or DAQ. "But you want to be absolutely sure that a positron, which is the primary thing we look for, is isolated from anything to do with cosmic rays. You need some way to separate the wheat from the chaff, so to speak. The only way you can convincingly sort it all out is by keeping track of what happened before the positron showed up and after it was detected."

The DAQ keeps track. It is an array of electronics (one circuit board is shown in the photo) that can be thought of as a kind of brain with the single-minded task of identifying potential neutrino events whose signature is a positron followed by a 2.2-MeV gamma ray. When the DAQ identifies a promising positron, it grabs from its short-term memory everything that happened in the tank for 6 microseconds (μs) before the positron was detected. It also records all gamma-like activity that occurs within the next 1 millisecond (ms). Armed with this information, the DAQ tries to make sense of what it saw by correlating, in space, time, and energy, the positron signal with a gamma ray. If the correlation matches the profile of a neutrino interaction, it writes the information to long-term memory (a magnetic tape).

The human brain, with its exceptional pattern recognition ability, evolved over hundreds of millions of years. The DAQ used by LSND was designed and assembled in less than 1 year by Sandberg and a team of students and visiting staff as they scrambled for funding and raced to finish before LANSCE started producing neutrinos. By all accounts, the final system has been a smashing success. "It's a unique system," says Darryl Smith, who was involved in developing the DAQ. "No other data collection system has this look-back capability, to see what was happening in the detector *before* the triggering signal occurred. If in the future someone asks, 'Did you check for this, or look for this oddball correlation,' we can go back to the data and see."

How the DAQ searches for a neutrino signature is shown in the diagram on pages 114 and 115. In the text that follows, the letter callouts correlate with those on the figure.

The Front End—Selecting Promising Signals. (A) The DAQ perceives the world through 1,220 photomultiplier tubes (PMTs) that line the inside of the detector tank. Like huge eyes, the phototubes watch for the brief pulses of light produced when energetic particles pass through the oil. Outside the tank, 292 more phototubes sit within the cosmic-ray veto shield to signal the arrival of cosmic rays. When any tube is hit by a photon, it sends a tiny current pulse to a digital version of short-term memory called a circular buffer. There is one buffer per PMT.

(B) Conceptually, the circular buffer is analogous to an office Rolodex filled with 2,047 electronic "cards." At every time T, where T is the system time [measured in 100-nanosecond (ns) units, that is, $(T) - (T-1) = 100$ ns], the Rolodex is "turned" and the value of the electric charge q in the current pulse and the precise time t that the pulse occurred (accurate to about ±0.5 ns) are written to a card. The card is stamped with the system time T. If a tube was not hit, $q = 0$ and $t = 0$, but the data are still written to the card, which still receives a time stamp. Because a tube is hit on average only once every 200 μs, the data on most cards are just zeroes.

At 100 ns per card and 2,047 cards, each circular buffer maintains a 204-μs history of its phototube. The DAQ can access any portion of that history by asking the buffer to "dump" the

This circuit board is one of 210 that make up the LSND data acquisition system. A board contains eight individual channels, and each channel consists of analog circuitry for processing signals from one photomultiplier tube and digital circuitry that stores and/or sends that data to the computer. The board is built from mostly off-the-shelf components and is thus fairly robust and inexpensive. To achieve a flexible yet highly reliable system, each board was designed to be used in a standard VME electronics crate. This means that the DAQ can be easily adapted to meet the needs of other experiments.

information contained on specific cards, for example, the cards stamped $T-1$ through $T-60$. Because it has access to data from all 1,512 circular buffers, the DAQ in principle can construct detailed, 204-μs histories of everything that occurred in the tank and veto shield.

But a full history comes at a steep price tag. There are over 15 million pieces of data coming in per second, far too much data for a computer to mull over in detail. The system needs to quickly cull promising positron signals from the heavy traffic of cosmic rays and other "stuff" that cruises through the tank. Thus, the DAQ applies some rules of thumb so that it can react "instinctively." (*C*) In addition to producing an analog current pulse, each phototube when hit by a photon sends a digital pulse to a summing circuit. The digital pulse simply indicates that the tube fired in the preceding 100 ns, and the sum of the pulses gives the total number of tubes that fired in the tank and the veto shield within that time interval. (Tank and veto shield sums are kept separate.) The information goes to a programmable "trigger" (called the signal flagger) that crudely identifies the signal. For example, if at least 21 phototubes fire in the detector and less than 4 fire in the veto shield, the signal is flagged as a gamma-ray candidate. If at least 150 tubes fire in the detector and less than 4 fire in the shield, the signal is flagged as a positron candidate. Cosmic-ray muons typically light up the tank like a Roman candle, setting off 250 to 1,000 or more tubes. Flagged signals, and the times that they occurred, are passed onto a trigger computer charged with the task of selecting promising ones.
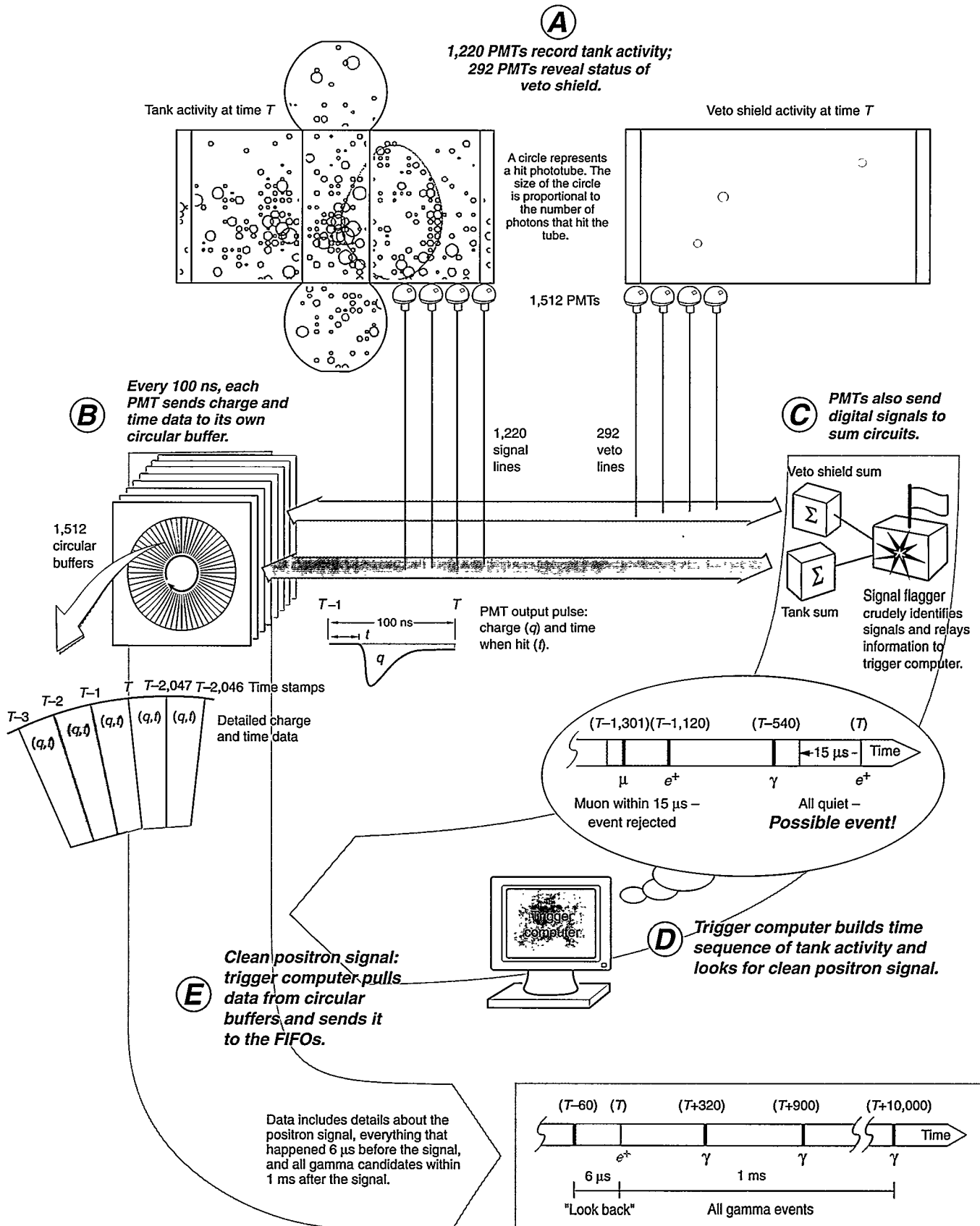
(*D*) The trigger computer monitors and stores signals as they roll in, always on the lookout for a positron candidate. When it sees one, it reviews the signal roster, checking back in time to see if there was any activity in the tank within a 15-μs "all-quiet" period before the positron appeared. If there was activity, it ignores the signal and keeps looking. The reason for the all-quiet condition is to weed out positrons that come from muons.

"Cosmic-ray muons that decay in the tank are the bugaboo of this experiment," says Sandberg. "They produce a Michel electron that is identical to the positron produced by electron antineutrinos. The average lifetime of the muon is about 2 μs, so if there's any suspicious activity within about 7 muon lifetimes before a positron signal, we don't want to waste time checking that positron. We're not interested."

Cosmic rays, beta particles from radioactive decay, neutrons, and muons create a riotous background of activity in the tank, and flagged signals are sent to the trigger computer at an average rate of one every 60 μs or so. Positrons account for about a third of that rate. With the all-quiet condition, the rate of signals that will be examined further is cut down to about one every 10 ms. That rate is slow enough that a large, mainframe "analysis" computer—the cerebrum of the DAQ brain—can examine the selected signals in detail. Because the decision to flag a signal is based on low-level information (the total number of tubes that fired in the tank and veto shield), it can be made quickly. (*E*) Within about 400 ns of finding a clean positron with no prior history, the trigger computer sends a message to all circular buffers to dump their detailed information.

The Back End—Reconstructing Events. (*F*) If the positron signal occurred at time T, the trigger computer tells the circular buffer to dump the data stored on the card stamped T into a temporary storage bin called a FIFO (an acronym for first in, first out). There is one FIFO per photomultiplier. In addition, the trigger tells the circular buffer to dump into the FIFO the 60 pieces of (q,t) data corresponding to the 6 μs before T. This "look-back" information is a double check on the signal. Anything that happened in the 6 μs before the positron made itself known in the tank will be looked at in detail. Finally, all gamma-like signals that occurred within 1 ms after T are also placed in the FIFO.

The FIFO allows the analysis computer, which has the tough job of trying to figure out what happened, to leisurely collect the detailed information from each tube. (*G*) The analysis

A 1,220 PMTs record tank activity;
292 PMTs reveal status of
veto shield.

Tank activity at time *T*

Veto shield activity at time *T*

A circle represents
a hit phototube. The
size of the circle
is proportional to
the number of
photons that hit the
tube.

1,512 PMTs

1,220
signal
lines

292
veto
lines

B Every 100 ns, each
PMT sends charge and
time data to its own
circular buffer.

C PMTs also send
digital signals to
sum circuits.

Veto shield sum

Σ

Σ

Tank sum

Signal flagger
crudely identifies
signals and relays
information to
trigger computer.

1,512
circular
buffers

$T-1$ T

$\leftarrow$ 100 ns $\rightarrow$

t

PMT output pulse:
charge (q) and time
when hit (t).

q

$T-3$ $T-2$ $T-1$ T $T-2,047$ $T-2,046$ Time stamps

(q,t) (q,t) (q,t) (q,t) (q,t) Detailed charge
and time data

$(T-1,301)(T-1,120)$ $(T-540)$ (T)

$\leftarrow$15 μs $\rightarrow$ | Time

μ e^+ γ e^+

Muon within 15 μs –
event rejected

All quiet –
Possible event!

E Clean positron signal:
trigger computer pulls
data from circular
buffers and sends it
to the FIFOs.

D Trigger computer builds time
sequence of tank activity and
looks for clean positron signal.

trigger
computer

Data includes details about the
positron signal, everything that
happened 6 μs before the signal,
and all gamma candidates within
1 ms after the signal.

$(T-60)$ (T) $(T+320)$ $(T+900)$ $(T+10,000)$

Time

e^+ γ γ γ

6 μs   1 ms

"Look back"   All gamma events

Electron antineutrino
Positron
Neutron

H *The reconstructed neutrino event is written to tape: a positron with the right energy at (x,y,z), followed in time by a gamma ray with the right energy at (x',y',z').*

Neutrino event!

(T) (T+320) (T+770) (T+1,411) (T+1,890)

Time

e^+ γ γ γ γ

$R > 30$

Reconstructed signals

G *Analysis computer reconstructs signals, looking for position and time correlations between gamma rays and positrons.*

F *FIFOs temporarily store data.*

1,512 individual FIFOs

T–3 T–2 T–1 T T+320

(q,t) (q,t)

Everything   Gamma events
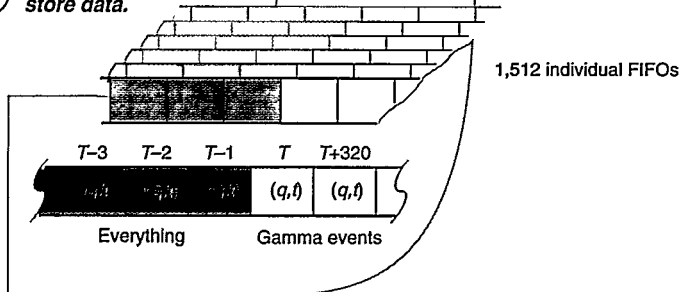
computer gathers the detailed photomultiplier data from all 1,512 FIFOs and reconstructs the signals. First, it finds the Cerenkov cone and the sphere of scintillation light and determines the positron's trajectory and position. Next, it finds the position and energy of the gamma rays (by reconstructing their spheres of scintillation light) and then uses the likelihood function R to find a gamma that has the right time, energy, and position to have come from neutron capture. If R is high (>30), the sequence of signals—a positron followed by a gamma ray from neutron capture—is taken as the signature of a neutrino event. (**H**) The reconstructed neutrino events are written to tape.

Since 1993, the trigger computer has looked at half a billion flagged signals, of which twenty-two were identified as the signature of electron antineutrinos.[*] It was a significant challenge to design a system that could handle such a low, "asynchronous" event rate. The DAQs used in most particle physics experiments operate on a clock that is synchronized with the particle beam, so that the electronics know exactly when to pay attention. By contrast, neutrino oscillation events in the LSND experiment appear at almost random times. The DAQ has to operate continuously, look at all events, but select only a tiny subset. "Traditional experiments threw stuff away," says Sandberg. "We couldn't afford to do that. But we also couldn't afford to keep it all." The front-end selection of signals and the look-back capability of the DAQ helped solve the problem.

"Because neutrino oscillations have come and gone from decade to decade," says Smith, "we needed as much credibility as we could bring to bear on the problem. This DAQ is totally solid. It's built almost entirely from off-the-shelf components. It can live in an enormously noisy environment. It can look back in time. Because it's built using a standard VME architecture, any graduate student could plunk it into an experiment and get it up and running. It works." ■

[*] Through a separate analysis, nineteen electron neutrino events were also identified.

The Evidence
for Oscillations

Bill Louis, Vern Sandberg, Gerry Garvey,
Hywel White, Geoffrey Mills, and Rex Tayloe
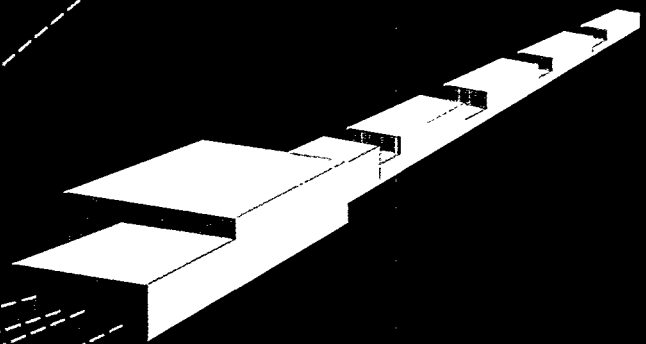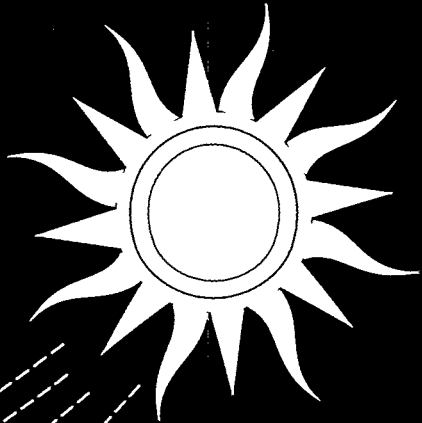
In the Standard Model of particle physics, the masses of quarks and the mixing between quark mass states are well known. For leptons, however, the neutrino masses are unknown (and it is even questionable whether neutrinos have mass). If anything, neutrinos weigh very little—current mass limits are $\nu_e < 10$ electron volts[1] (eV), $\nu_\mu < 170$ kilo-electron-volts (keV), $\nu_\tau < 24$ million electron volts (MeV)—but exactly how little is something physicists would like to determine. A neutrino mass of only a few electron volts, for example, would likely affect cosmology and possibly affect the evolution of the universe.

Neutrino oscillations offer one of the best ways to measure small neutrino masses and mixings. (For the purists, oscillations also represent a beautiful example of quantum mechanics.) Oscillations refer to a periodic changing of one neutrino type into another, a phenomena that can occur only if neutrinos have mass. In that case, neutrinos would be described by three states ν_1, ν_2, and ν_3 with masses m_1, m_2, and m_3, respectively. If our understanding of the quarks is to guide our thinking about the leptons, however, those mass states would be different from the states associated with weak decays (the flavor states ν_e, ν_μ, and ν_τ). The flavor states would most likely be mixtures of the mass states. (Mass and mixing are discussed in detail in the primer, "The Oscillating Neutrino," on page 28.)

Consider, for example, a model in which only two neutrino types mix together (two-generation mixing). The electron neutrino and the muon neutrino are conventionally described as a combination of ν_1 and ν_2:

$$\nu_e = \cos\theta\, \nu_1 + \sin\theta\, \nu_2$$
$$\nu_\mu = -\sin\theta\, \nu_1 + \cos\theta\, \nu_2 \ .$$

The angle θ is called the *mixing angle*. It is an arbitrary parameter that can be determined only by experiment. Note that if θ is small, there is an approximate one-to-one correspondence between the flavor states and the mass states, that is, $\nu_e \approx \nu_1$ and $\nu_\mu \approx \nu_2$.

The fascinating aspect about mixing is what it implies for the neutrinos. Once born, a muon neutrino has some probability of being detected as an electron neutrino. That probability depends on the distance x that the muon neutrino has traveled and is given by the expression

$$P(\nu_\mu \to \nu_e) = \sin^2 2\theta \, \sin^2\left(\frac{1.27\Delta m^2 x}{E_\nu}\right) \ ,$$

where $\Delta m^2 = m^2_2 - m^2_1$ (the difference of the squares of the neutrino masses) in electron volts squared (eV^2), E_ν is the neutrino energy in million electron volts, and x is measured in meters. The expression is essentially that of a sinusoidal wave,

$$P(\nu_\mu \to \nu_e) = A \, \sin^2\left(\frac{\pi x}{\lambda}\right) \ ,$$

with the amplitude of the wave given by $A = \sin^2 2\theta$, and the wavelength, λ, which is also called the oscillation length, given by $\lambda = \pi E_\nu / 1.27\Delta m^2$. The mass difference Δm^2 can be determined from the oscillation wavelength, while the mixing angle θ is deduced from the wave's amplitude.

There is every reason to believe, however, that each flavor neutrino would be a mixture of all three mass states (three-generation mixing). The formalism for three-state mixing is a little more complex and yields an expression for the oscillation probability that is similarly more complex than the simple expression given above. For example, instead of a single parameter θ characterizing the mixing, there are three independent parameters. Likewise, there are three mass differences, (although only two are independent). Fortunately, if the mass scales are quite

Table I. Solar-Neutrino Data: Contributions to the Detected Signal from Each Solar-Neutrino-Producing Reaction (Expressed as a Fraction of the Total Signal) and the Ratio of the Measured Rate to the Predicted Rate

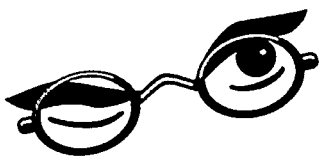|	SAGE and GALLEX	Chlorine	Kamiokande
Reaction			
pp	0.538		
^{7}Be I	0.009		
^{7}Be II	0.264	0.150	
^{8}B	0.105	0.775	1
pep	0.024	0.025	
CNO	0.060	0.041	
Ratio*	0.62 ± 0.1	0.29 ± 0.03	0.29 ± 0.03

*Predicted rate based on the Bahcall-Pinsonneault standard solar model.

different ($m_3 \gg m_2 \gg m_1$, for example), then the two-generation mixing model is a good approximation to the three-generation model. Therefore, the data from any one experiment is often analyzed in terms of the two-generation oscillation parameters Δm^2 and $\sin^2 2\theta$. It should be kept in mind, however, that such an analysis is likely an oversimplification of the physics, and different experiments may indicate very different mass differences and mixing angles for the same type of oscillation.

The potential for a given neutrino type to appear or disappear suggests two ways to perform neutrino oscillation experiments: either one looks for a *decrease* in the flux of the original neutrino type (known as a disappearance experiment), or one looks for the presence of a neutrino type other than the original one (an appearance experiment). Ideally, one also measures the neutrino flux as a function of distance, so that periodic variations of the flux become apparent.

Because neutrino detectors are so large, moving one back and forth and from place to place is essentially impossible. Instead, "single-point" measurements are made at a fixed distance from the source, which has the unpleasant consequence of making the search for oscillations a shot in the dark. The ideal distance between a neutrino source and detector would be half of the oscillation length, but that length cannot be calculated in advance because the mass difference Δm^2 is unknown. If the detector happens to be placed well within a quarter of an oscillation length from the source, most neutrinos will be in their original state and no effect will be seen. Likewise, if the detector happens to be many oscillation lengths away, then the unavoidable spread in the momenta of the original neutrinos will wash out any oscillation effects. An experiment is sensitive to only a range of oscillation lengths and will probe only a limited set of possible values of the neutrino mass difference and mixing angles.

Despite these limitations, more than a dozen experiments, spanning three decades of research, have been mounted in an effort to observe the chance metamorphosis of one neutrino type into another. To date, there is no smoking gun in the experimental evidence, only tantalizing hints that such a startling transformation does indeed occur. Without conclusive proof, physicists argue vehemently over whether neutrino oscillations have been observed. The debate is entirely appropriate. On an issue as important as the initial evidence for mixing in the lepton sector and neutrino mass, the scientific community requires convincing proof before accepting any claims. One facet to the debate that should be empha-

sized is that the evidence for oscillations comes from three different neutrino sources: neutrinos born in the core of the Sun, neutrinos created in the earth's upper atmosphere, and neutrinos that emanate from man-made sources (accelerators and reactors). Each of these sources provides different fluxes, energies, and types of neutrinos. They also have different sources of errors. Disappearance experiments that look at solar or atmospheric neutrinos have all indicated a reduced neutrino flux, but the reduction is only relative to a theoretical prediction. Because the neutrinos being observed are created by natural processes and hence are not completely characterized, physicists cannot rule out that the apparent reduction in flux is merely a theoretical miscalculation. The neutrinos streaming out of man-made sources are better characterized, but the experiments must contend with much greater backgrounds. At present, only one appearance experiment—the accelerator-based liquid scintillator neutrino detector (LSND) experiment—has seen evidence for oscillations.

Experimental Evidence

Evidence from Solar Neutrinos. Since the first observation of electron neutrino interactions in a chlorine-laden tank by Ray Davis, Jr., and collaborators, three additional experiments have measured solar-neutrino interactions. Two of those, SAGE and GALLEX, used gallium as a neutrino target, while the third, Kamiokande, used a water-Cerenkov detector in which neutrinos undergo elastic scattering with electrons in the water. All of the experiments have determined that there are fewer neutrinos from the Sun than are expected from the Standard Solar Model (see the article "Exorcising Ghosts" on page 136.)

The four experiments are sensitive to different parts of the solar-neutrino spectrum. The two gallium experiments have the lowest energy threshold (0.233 MeV) and are sensitive to the entire solar-neutrino flux. The chlorine experiment is sensitive to neutrinos with energies greater than 0.8 MeV, while Kamiokande is limited to detecting neutrinos with energies greater than about 7 MeV. The sensitivity of each experiment to the Sun's neutrino-producing reactions and the experimental results are listed in Table I. It is worth emphasizing that all the data presented in the table were gathered over an extended period of time and that each of the experiments has undergone numerous systematic checks. The solar models used to predict the neutrino flux are very much constrained by measured physical parameters, such as the solar luminosity, and like the experiments themselves, have been laboriously tested and refined over the years.

In a 1994 paper, Hata and Langacker (1994) presented a thorough analysis of all of the solar-neutrino data. They considered experimental errors in detail as well as possible variations to the standard solar model. They concluded that the experimental data cannot be explained by variations in solar physics and that neutrino oscillations are strongly favored. Furthermore, the most promising solution is a matter-enhanced, resonant transformation of electron neutrinos to other flavors through the MSW effect (see the article "MSW" on page 156). Given the range of densities in the Sun, the MSW effect could occur if the oscillation length λ is in the range of 10^4 to 10^8 meters. MSW leads to the allowed regions in the Δm^2 and $\sin^2 2\theta$ parameter space shown in Figure 1. The better fit to the data is obtained with a mass difference (Δm^2) of $\approx 10^{-5}$ eV^2 and the smaller mixing angle leading to a $\sin^2 2\theta$ value of $\approx 3 \times 10^{-3}$.

Overall, the experimental evidence that solar neutrinos may undergo oscillations appears firm, although uncertainties in solar dynamics are still a cause for concern.
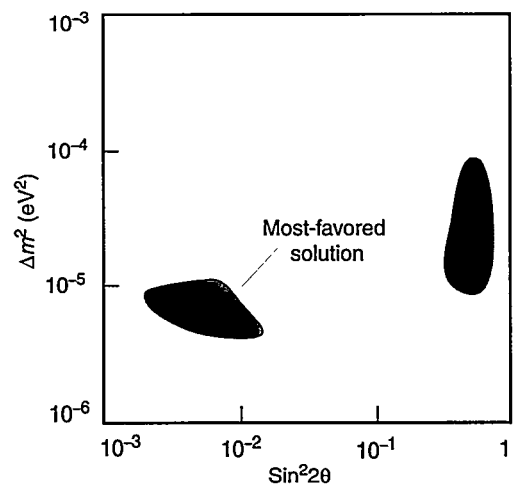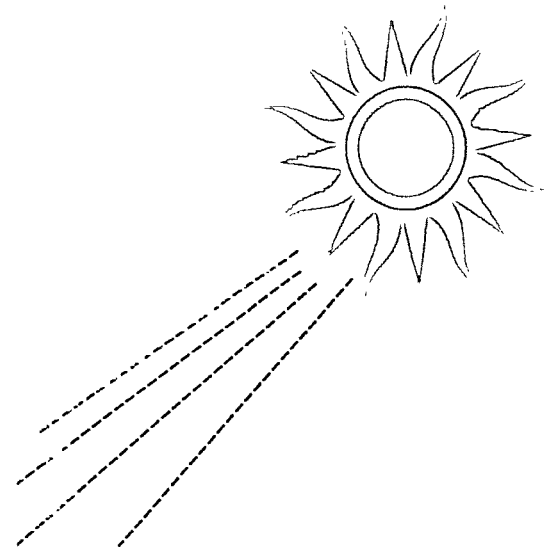
Figure 1. MSW Solutions to the Solar-Neutrino Problem
Given oscillations between only two neutrino mass states (two-generation mixing), a plot can be constructed that shows the values of Δm^2 and $\sin^2 2\theta$ that yield an oscillation probability consistent with the solar-neutrino data. The MSW effect can occur within a continuous range of mass differences and mixing angles, but the four experiments exclude certain values. Two regions (the MSW solutions) are allowed by all four experiments. The region to the left, with the smaller mixing angles, is the one most favored by theorists.

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$$

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu$$

Cosmic ray

$$\pi^+ \rightarrow \mu^+ + \nu_\mu$$

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu$$

π^-

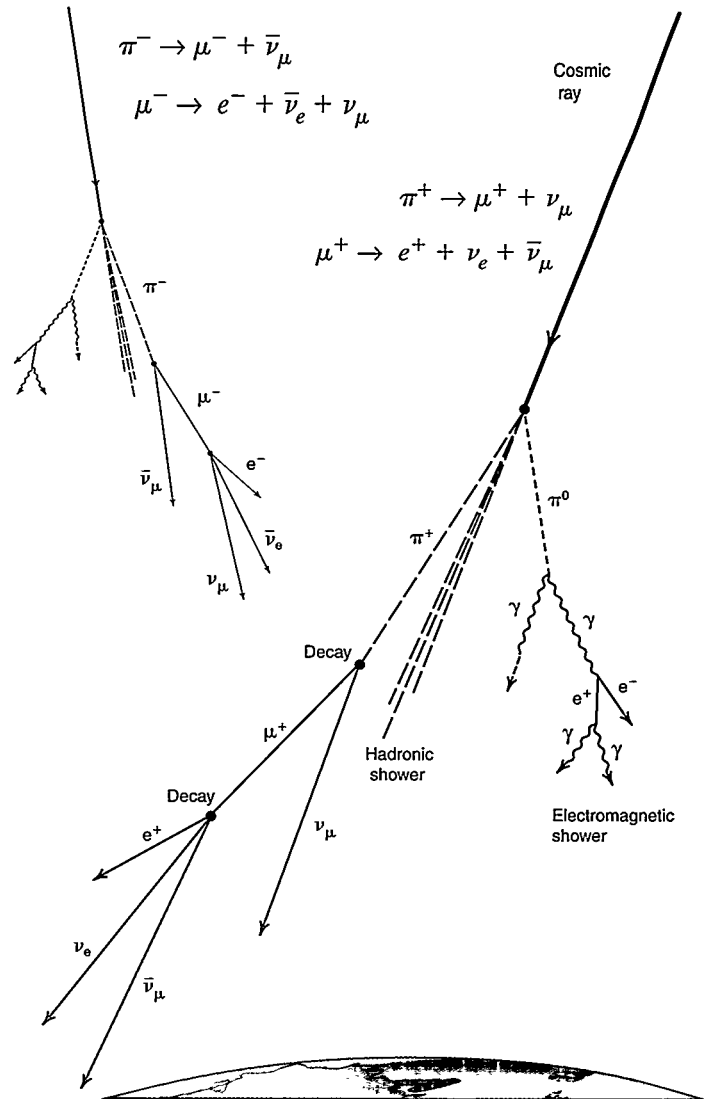μ^-

e^-

$\bar{\nu}_\mu$

$\bar{\nu}_e$

ν_μ

π^0

π^+

Decay

μ^+

Hadronic shower

Decay

e^+

ν_μ

γ

γ

e^+ e^-

γ γ

Electromagnetic shower

ν_e

$\bar{\nu}_\mu$

Figure 2. The Atmospheric-Neutrino Source

Collisions between cosmic rays and nuclei in the upper atmosphere can create high-energy pions (π). In the collision shown on the right, a π^+, π^0, and other heavy particles (the hadronic shower) are created. The π^0 decays and produces gamma rays and leptons (the electromagnetic shower) but no neutrinos. The π^+ produces two muon neutrinos (blue) and an electron neutrino (red). The collision shown on the left produces a π^-, leading to the production of two muon neutrinos and an electron antineutrino.

The result of the Kamiokande experiment will be tested in the near future by super-Kamiokande, which will have significantly better statistical precision. Also, the neutrino oscillation hypothesis and the MSW solution will be tested by the Sudbury Neutrino Observatory (SNO) experiment, which will measure both charged- and neutral-current solar-neutrino interactions.

Evidence from Atmospheric Neutrinos. Upon reaching the earth, high-energy cosmic rays collide violently with nuclei present in the rarefied gas of the earth's upper atmosphere. As a result, a large number of pions—π^-, π^0, and π^+—are produced (see Figure 2). These particles eventually decay into either electrons or positrons and various types of neutrinos and antineutrinos. (A large number of kaons are also produced by cosmic rays in the upper atmosphere, and these particles also eventually decay into various leptons.) As seen in Figure 2, the decay of either positive or negative pions results in the eventual production of two muon neutrinos (ν_μ and $\bar{\nu}_\mu$) but only one electron neutrino (either ν_e or $\bar{\nu}_e$). Experimenters, therefore, expect to measure two muon neutrinos for each electron neutrino.

Atmospheric neutrinos are orders of magnitude less abundant than solar neutrinos, but can be readily detected because they have very high energies.

Table II. Results from the Atmospheric Neutrino Experiments

Experiment	Exposure (kiloton-year)	R
IMB I	3.8	0.68 ± 0.08
Kamiokande Ring	7.7	0.60 ± 0.06
Kamiokande Decay	–	0.69 ± 0.06
IMB III Ring	7.7	0.54 ± 0.05
IMB III Decay	–	0.64 ± 0.07
Frejus Contained	2.0	0.87 ± 0.13
Soudan	1.0	0.64 ± 0.19
NUSEX	0.5	0.99 ± 0.29

(The neutrino interaction cross sections, and hence the neutrino detection probability, increases dramatically with energy.) Depending on the energy of the incident cosmic ray and how its energy is shared among the fragments of the initial reaction, neutrino energies can range from hundreds of millions of electron volts to about 100 giga-electron-volts (GeV). (In comparison, the highest-energy solar neutrino comes from the ^8B reaction, with a maximum energy of about 15 MeV.)

Muon neutrinos produce muons in the detector, and electron neutrinos produce electrons, so that the detector signals can be analyzed to distinguish muon events from electron events. Because the sensitivity of the detectors to electrons and muons varies over the observed energy range, the experiments depend on a Monte Carlo simulation to determine the relative detection efficiencies. Experimental results, therefore, are reported as a "ratio of ratios"—the ratio of observed muon neutrino to electron neutrino events divided by the ratio of muon neutrino to electron neutrino events as derived from a simulation:

$$R = \frac{(\nu_\mu/\nu_e)_{\text{observed}}}{(\nu_\mu/\nu_e)_{\text{simulation}}}$$

If the measured results agree with the theoretical predictions, $R = 1$.

A recent summary of the experimental data is given by Gaisser and Goodman (1994) and shown in Table II. For most of the experiments, R is significantly less than 1: the mean value is about 0.65. (In the table, the Kamiokande and IMB III experiments identify muons in two ways. The first involves identification of the Cerenkov ring, which is significantly different for electrons and muons. The second involves searching for the energetic electron that is the signature for muons that have stopped in the water detector and decayed. A consistent value of R is obtained using either method.) Despite lingering questions concerning the simulations and some systematic effects, the experimenters and many other physicists believe that the observed values for R are suppressed by about 35 percent.

The Kamiokande group has also reported what is known as a zenith-angle dependence to the apparent atmospheric-neutrino deficit. Restricting the data to neutrinos that come from directly over the detector (a zenith angle of 0 degrees and a distance of about 30 kilometers) yields $R < 1.3$ (that is, more muon to electron neutrino events are observed than predicted by theory). Neutrinos that are born closer to the horizon (a zenith angle of 90 degrees) and have to travel a greater distance to reach the detector result in $R < 0.5$. Finally, neutrinos that have to travel through the earth to reach the detector (roughly 12,000 kilometers) result in an even lower value for R. The apparent
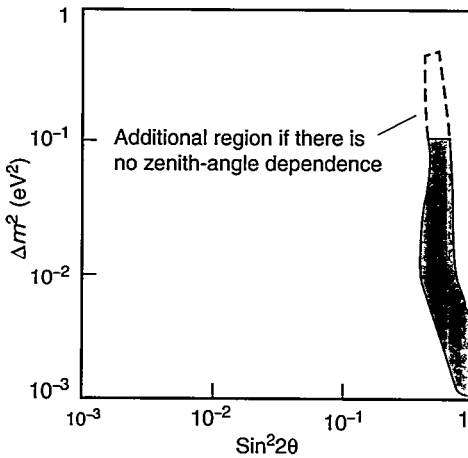
zenith-angle dependence shows up only for neutrinos with energies greater than 1.3 GeV.

This single piece of evidence has had a significant impact on the allowed region of Δm^2 and $\sin^2 2\theta$ (see Figure 3). The fact that little disappearance effect is observed for a zenith angle of ~0 degrees means that the oscillation length is much greater than 30 kilometers, so that

$$\frac{\pi E_\nu}{1.27\ \Delta m^2} \gg 30\ .$$

With $E_\nu \approx 6$ GeV, one finds that $\Delta m^2 \ll 0.5$ eV^2. Given this small value for Δm^2, neutrinos emerging from some high-energy accelerators would have oscillation lengths on the order of hundreds of kilometers. A number of proposals have suggested placing huge neutrino detectors at comparable distances from an accelerator in an effort to investigate $\nu_\mu \rightarrow \nu_e$ oscillations.

However, the statistical significance of the reported zenith-angle dependence is not large. Moreover, a preprint from the Irvine-Michigan-Brookhaven (IMB) collaboration reports no such dependence, and early data from the experiment that has succeeded Kamiokande—super-Kamiokande—is consistent with only a slight zenith-angle dependence. If the zenith-angle dependence disappears, then the atmospheric data is consistent with $\Delta m^2 > 0.15$ eV^2 and an oscillation length on the order of 20 kilometers. This value of Δm^2 is compatible with the LSND observation discussed below.

Evidence from Accelerator-Produced Neutrinos. To date, LSND at the Los Alamos Neutron Science Center (LANSCE) is the only accelerator experiment to have evidence for neutrino oscillations. The experiment uses a detector that contains 167 metric tons of dilute liquid scintillator placed 30 meters from the beam stop for the LANSCE proton beam. Neutrinos are produced from the decay of positive pions that come to rest in the beam stop:

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \quad \text{and} \quad \mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu\ .$$
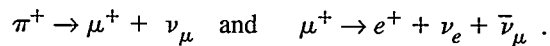
No electron antineutrinos are produced in this reaction chain. Thus, LSND seeks evidence for $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ oscillations by looking for electron antineutrino interactions in the detector. The charged-current weak interaction of electron antineutrinos with protons results in the creation of a positron and a free neutron:

$$\bar{\nu}_e + p \rightarrow e^+ + n\ .$$
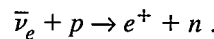
The positron instantly streaks through the detector and produces both Cerenkov and scintillation light. The neutron, after a mean lifetime of 186 microseconds, is captured by a proton to form deuterium, D, and a 2.2-MeV gamma ray is produced:

$$n + p \rightarrow D + \gamma\ .$$
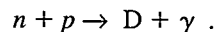
The gamma ray also creates light in the detector. The signature for the electron antineutrino event is the correlation of the positron's Cerenkov and scintillation light with scintillation light produced by the 2.2-MeV gamma ray. Because of the low energy of the LANSCE beam (800 MeV), the neutrino backgrounds in LSND are quite small and well understood. The largest background is from electron antineutrinos that are produced when negative muons decay at rest in the beam stop. This decay channel, however, is suppressed by a factor of 7×10^{-4}

Figure 3. Allowed Δm^2 and $\sin^2 2\theta$ Region for Atmospheric Neutrinos

The data for the atmospheric-neutrino experiments is best explained with large mixing angles, and to be consistent with other experiments, the parameter space is limited to the region shown in gray. The disputed zenith-angle dependence of the data limits the values of Δm^2 to less than 0.1 eV^2. If the zenith-angle dependence can be discounted, then mass differences as high as 0.5 eV^2 are allowed. New data from the next generation of atmospheric-neutrino experiments should settle the question of zenith-angle dependence.
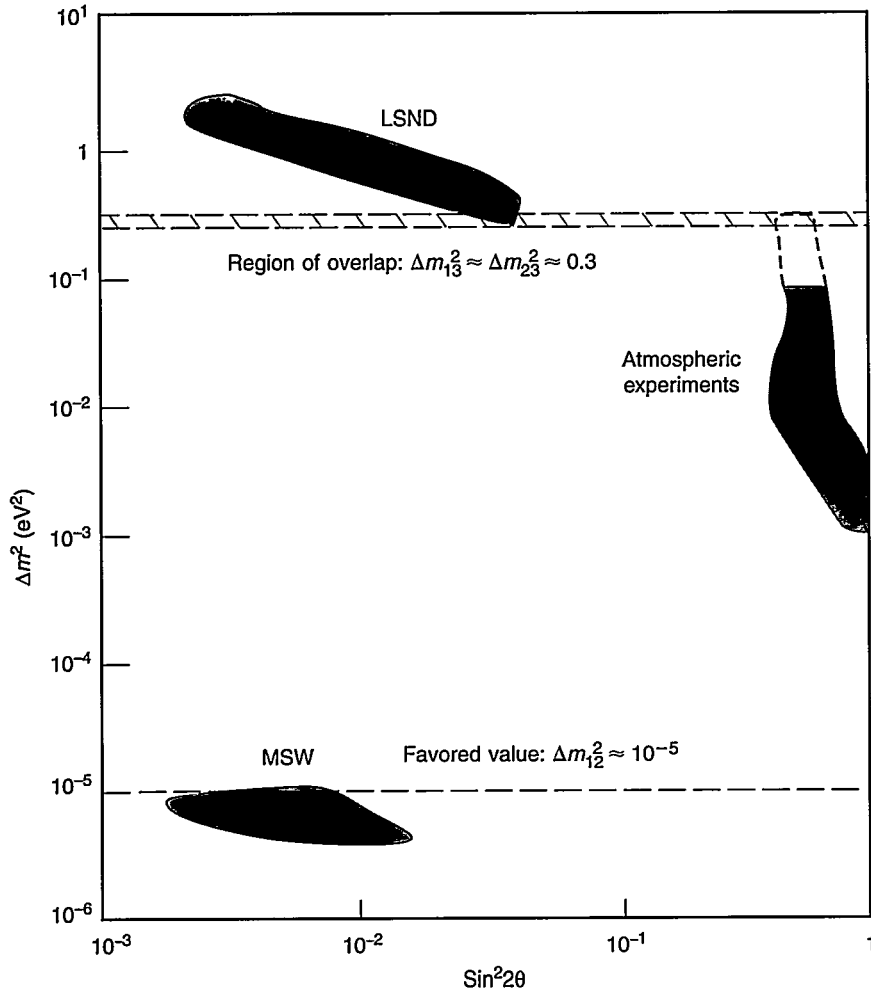
Figure 4. Allowed Regions for All Three Neutrino Sources
There are a minimum of two mass difference scales: one associated with the MSW solution to the solar-neutrino experiments ($\Delta m^2 \approx 10^{-5}$ eV^2) and the other associated with the atmospheric and LSND experiments ($\Delta m^2 \approx 10^{-1}$ eV^2), assuming no zenith-angle dependence. A three-generation mixing model is needed to explain all the data.

relative to the positive muons that decay at rest. The complete story of the LSND experiment is told in the article "A Thousand Eyes" on page 92.

The experiment has reported evidence for $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ oscillations by observing an excess of 22 electron antineutrino events above background. This number corresponds to an oscillation probability of $P(\bar{\nu}_\mu \rightarrow \bar{\nu}_e) \approx 0.3$ percent, with the Δm^2 and sin$^2 2\theta$ parameter region shown in Figure 4. For comparison, the figure also shows the regions of Δm^2 and sin$^2 2\theta$ allowed by the solar- and atmospheric-neutrino experiments.

The evidence for neutrino oscillations from LSND is strengthened by results from a complementary $\nu_\mu \rightarrow \nu_e$ oscillation search using the same detector and source. In the search, experimenters have observed an excess of 19 electron neutrino events above background. This "second" experiment (the two neutrino searches are actually performed simultaneously) has completely different systematic errors and backgrounds from those of the $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ oscillation search. The second set of neutrinos, which come from pions that decay in flight, have higher energies than those produced by muons that decay at rest. Thus, it is interesting that the decay-in-flight analysis shows a signal, although of lesser significance, that indicates the same favored regions of Δm^2 and sin$^2 2\theta$ as the decay-at-rest analysis. A comprehensive analysis of the decay-in-flight and decay-at-rest data is in progress.

Theoretical Interpretation of the Data

If one ignores the zenith-angle dependence of the atmospheric neutrinos, there appear to be two distinct mass differences implied by the data. As seen in Figure 4, one is a "small" mass difference associated with the solar neutrinos ($\Delta m^2 \approx 10^{-5}$ eV^2); the other is a "large" difference associated with the atmospheric and LSND experiments ($\Delta m^2 \approx 10^{-1}$ eV^2). The data is particularly puzzling with regard to the solar and LSND results. Both experiments presume to be observing oscillations between electron and muon neutrinos, but it takes very disparate mass differences to explain their respective data sets. It is therefore natural to ask whether any consistent picture can be made of all the experimental results. (If the zenith-angle dependence is shown to be valid, then there are three distinct mass differences and the answer is no: the data cannot be explained by any consistent oscillation formalism involving only three neutrinos.)

As stated at the beginning of this article, analysis of the data in terms of a two-generation mixing model may be an oversimplification of the physics. In that model, only two neutrino types are considered. Rewriting the equations for electron neutrinos and muon neutrinos in matrix form yields

$$\begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_{12} & \sin\theta_{12} \\ -\sin\theta_{12} & \cos\theta_{12} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} .$$

The 2×2 mixing matrix contains only sines and cosines and depends on a single parameter, θ_{12}. (The subscripts have been added to emphasize that θ_{12} characterizes the mixing between states ν_1 and ν_2.)

Mixing between three neutrino generations means that not only are there three mass differences, Δm_{12}^2, Δm_{13}^2, and Δm_{23}^2, where $\Delta m_{ij}^2 = m_j^2 - m_i^2$, but there are also three independent mixing parameters. A simple 3×3 mixing matrix U can be constructed by taking the product of three unitary matrices.[2]

$$U = U_{12}U_{13}U_{23} ,$$

where

$$U_{12} = \begin{pmatrix} \cos\theta_{12} & \sin\theta_{12} & 0 \\ -\sin\theta_{12} & \cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} ,$$
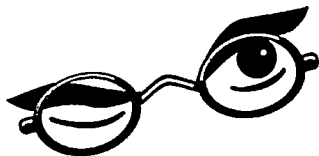
$$U_{13} = \begin{pmatrix} \cos\theta_{13} & 0 & \sin\theta_{13} \\ 0 & 1 & 0 \\ -\sin\theta_{13} & 0 & \cos\theta_{13} \end{pmatrix} , \text{ and}$$

$$U_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{23} & \sin\theta_{23} \\ 0 & -\sin\theta_{23} & \cos\theta_{23} \end{pmatrix} .$$

For convenience, the matrix U will be written as

$$U = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} .$$

The elements $U_{\alpha i}$, where $\alpha = e, \mu, \tau$ and $i = 1, 2, 3$, depend only on the products

[2] If there are more neutrino states than are allowed by the Standard Model (for example, if there are right-handed neutrinos), then the mixing matrix could be much larger than a 3×3 and could contain more than three parameters. The simple matrix presented here assumes no right-handed neutrinos and no CP-violating phase.

of the sines and cosines of the mixing angles θ_{12}, θ_{13}, and θ_{23}. Mixing between the three neutrino generations takes the form

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \ .$$

This formalism is analogous to the quark sector, where strong and weak states are not identical and the resultant mixing is described conventionally by a unitary mixing matrix (the Cabbibo-Kobayashi-Maskawa matrix).

Given the arbitrary mixing matrix above, the oscillation probability is

$$P(\nu_\alpha \to \nu_\beta) = \delta_{\alpha\beta} - 4 \sum_{j>i} U_{\alpha i} U_{\beta i} U^*_{\alpha j} U^*_{\beta j} \sin^2\left(\frac{\pi x}{\lambda_{ij}}\right) \ .$$

where

$$\lambda_{ij} = \frac{\pi E_\nu}{1.27 \Delta m^2_{ij}}$$

Just as in the two-generation case, the oscillation length depends upon the mass difference (in electron volts squared), the length x from the source (in meters), and the neutrino energy E_ν (in million electron volts). The oscillation amplitude depends upon the $U_{\alpha i}$.

In a three-generation formalism, an oscillation between two flavor neutrinos occurs through all three mass states. To be explicit, the oscillation probability between electron neutrinos and muon neutrinos is given by

$$\begin{aligned} P(\nu_e \to \nu_\mu) = \ & 4 U_{e1} U_{\mu 1} U_{e2} U_{\mu 2} \sin^2\left(\frac{\pi x}{\lambda_{12}}\right) \\ & + \ 4 U_{e1} U_{\mu 1} U_{e3} U_{\mu 3} \sin^2\left(\frac{\pi x}{\lambda_{13}}\right) \\ & + \ 4 U_{e2} U_{\mu 2} U_{e3} U_{\mu 3} \sin^2\left(\frac{\pi x}{\lambda_{23}}\right) \ . \end{aligned}$$

The first term in this expression (through λ_{12}) depends on the mass difference Δm^2_{12}. The second term depends on Δm^2_{13}, whereas the last depends on Δm^2_{23}. The coefficients in front of the sinusoidal terms involve all three mixing angles. Because there are multiple terms in the oscillation probability, $\nu_e \to \nu_\mu$ oscillations could appear to occur with different mass scales. An experiment could be sensitive to one or more oscillation lengths, depending on the specific source-to-detector distance x.

An example of a three-generation mixing model is the one put forth by Cardall and Fuller (1996). Their model ignores the zenith-angle dependence and sets $m_1 \approx m_2 \ll m_3$. All of the data from each of the three types of neutrino sources is then explained by the following mass differences and mixing matrix:

$$\Delta m^2_{12} \approx 10^{-5} \text{ eV}^2,$$
$$\Delta m^2_{13} \approx 0.3 \text{ eV}^2,$$
$$\Delta m^2_{23} \approx 0.3 \text{ eV}^2, \text{ and}$$

$$U_{\alpha i} \approx \begin{pmatrix} 0.99 & 0.03 & 0.03 \\ -0.03 & 0.71 & 0.71 \\ -0.03 & -0.71 & 0.71 \end{pmatrix} \ .$$

In the model used by Cardall and Fuller, electron neutrinos consist almost entirely
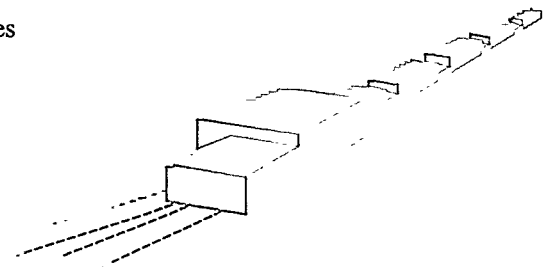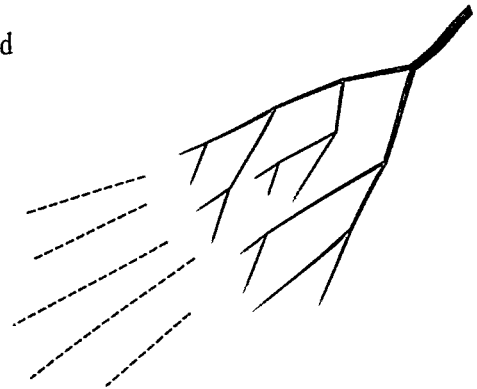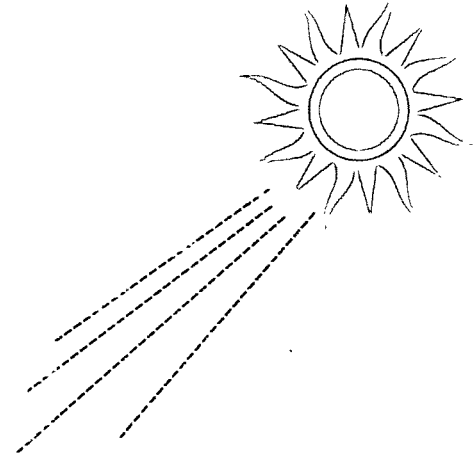
of the mass state ν_1, whereas muon and tau neutrinos are nearly identical particles that are mostly equal mixtures of the mass states ν_2 and ν_3.

The data from both LSND and the solar-neutrino experiments are explained as evidence for an oscillation between electron neutrinos and muon neutrinos, with an oscillation probability governed by $P(\nu_e \to \nu_\mu)$. Substituting the matrix elements $U_{\alpha i}$ into the formal equation for the oscillation probability yields

$$P(\nu_e \to \nu_\mu) \approx 0.0025 \, \sin^2\left(\frac{\pi x}{\lambda_{12}}\right) + 0.0025 \, \sin^2\left(\frac{\pi x}{\lambda_{13}}\right) + 0.0018 \, \sin^2\left(\frac{\pi x}{\lambda_{23}}\right).$$
$$\uparrow \qquad\qquad\qquad \uparrow \qquad\qquad\qquad \uparrow$$
$$\text{solar} \qquad\qquad\quad \text{LSND} \qquad\qquad \text{LSND}$$

For LSND, the distance between the detector and source, x, is approximately 30 meters. Given the mass differences and the neutrino energies, the probability is dominated by the last two terms, which means that LSND is observing "indirect" oscillations between muon neutrinos and electron neutrinos. Although those two neutrinos are most closely associated with m_1 and m_2, the oscillation occurs because of the mass difference between m_1 and m_3 (that is, Δm_{13}^2) and between m_2 and m_3 (Δm_{23}^2). (A "direct" oscillation would depend only on Δm_{12}^2.) This indirect oscillation has a negligible effect on the neutrinos coming from the Sun ($x \approx$ 140 million kilometers). The density of matter in the Sun, however, is such that the MSW effect can resonantly enhance oscillations between two neutrinos with a very small mass difference. Electron neutrinos oscillate into muon neutrinos as they travel from the core of the Sun to its surface (a distance on the order of 100,000 kilometers). The first term in the probability expression—the one that depends on Δm_{12}^2 through the term $\sin^2(\pi x/\lambda_{12})$—dominates in this case.

The atmospheric data is explained simply by having muon neutrinos oscillate into tau neutrinos, with a probability that is dominated by the last term in the expression for $P(\nu_\mu \to \nu_\tau)$:

$$P(\nu_\mu \to \nu_\tau) \approx 0.0018 \, \sin^2\left(\frac{\pi x}{\lambda_{12}}\right) + 0.0018 \, \sin^2\left(\frac{\pi x}{\lambda_{13}}\right) + 1.016 \, \sin^2\left(\frac{\pi x}{\lambda_{23}}\right).$$
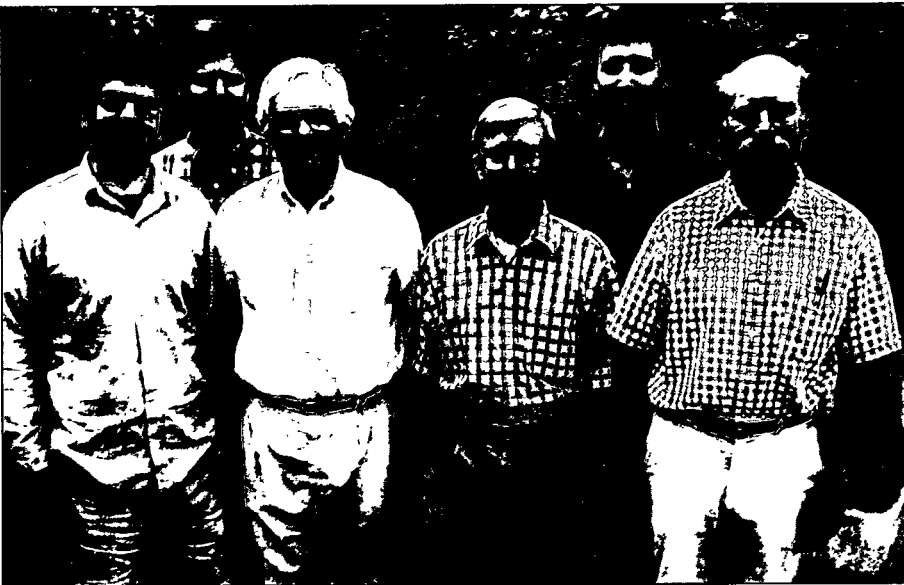$$\uparrow$$
$$\text{atmospheric}$$

Cardall and Fuller readily admit that their solution is "rather fragile," in that small adjustments to the allowed parameter space for any one of the experiments may not permit a global fit. Still, their solution currently explains all the data and sets up a "natural" framework for viewing the apparent disparity among experimental results. (It is important to emphasize that there are many possibilities other than the Cardall-Fuller solution. Some of these involve sterile neutrinos, inverted mass hierarchies, and new particles.)

Future Experiments

The possible solution by Cardall and Fuller will be tested in the near future by several ongoing and proposed experiments. Super-Kamiokande (located in the Kamioka Mine in Japan) and BOREXINO (located in the Gran Sasso tunnel in Italy) will test the results of the solar-neutrino experiments. SNO, which is located in the Creighton Nickel Mine in Canada, has the capability to measure neutral- and charged-current neutrino interactions. The experiment should directly test the solar-neutrino oscillation hypothesis and MSW solution and could possibly test for the existence of sterile neutrinos.

Super-Kamiokande will also continue to take data on atmospheric neutrinos. Establishing the statistical significance of any zenith-angle dependence is one of its major goals. CHORUS and NOMAD, two experiments at the European Center for Nuclear Research (CERN), are looking directly for muon neutrino to tau neutrino oscillations. There are also proposals to look for the appearance of tau neutrinos in a beam of muon neutrinos produced by a distant accelerator. MINOS would be located in the Soudan Mine in Minnesota and would use Fermilab in Illinois as its neutrino source. ICARUS, situated in the Gran Sasso tunnel, would be 732 kilometers from its neutrino source at CERN. ICARUS could also be sensitive to solar and atmospheric neutrinos.

KARMEN, located at the ISIS pulsed-neutron spallation source in Great Britain, and BOONE, to be located near Fermilab, will test the LSND solution. Together, these current and proposed experiments should be able to prove whether neutrino oscillations are indeed responsible for the discrepancies between theory and data. In the future, however, we look forward to the day when neutrino oscillation experiments move from the "discovery" of neutrino oscillations to the measurement of oscillation parameters, neutrino masses, and lepton-sector mixing angles. ■

Further Reading

Babu, K. S., et al. 1995. Indirect Neutrino Oscillations. *Physics Letters B* **359**: 351.

Cardall, C. Y., and G. M. Fuller. 1996. Can a Natural" Three-Generation Neutrino Mixing Scheme Satisfy Everything? *Physical Review D* **53**: 4421.

Fogli, G. L., E. Lisi, and G. Scioscia. 1995. Accelerator and Reactor Neutrino Oscillation Experiments in a Simple Three-Generation Framework. *Physical Review D* **52**: 5334.

Gaisser, T. K., and M. Goodman. 1994. *Proceedings of 1994 Snowmass Summer Study, Particle and Nuclear Astrophysics in the Next Millennium.* Edited by E. W. Kolb and R. D. Peccei. Singapore: World Scientific.

Hata, N., and P. Langacker. 1994. Model-Independent Determination of the Solar Neutrino Spectrum with and without the MSW Effect. *Physical Review D* **49**: 3622.

Left to right: Rex Tayloe, Geoffrey Mills, Hywel White, Vern Sandberg, Bill Louis, and Gerry Garvey. For biographies of Hywel White, Vern Sandberg, and Bill Louis, see "A Thousand Eyes." For a biography of Gerry Garvey, see "The Oscillating Neutrino."

Rex Tayloe was born and grew up in central Indiana. After earning his B.S. in physics from Purdue University, Tayloe attended the University of Illinois, where he received his Ph.D. for studies of hyperon physics at the Low Energy Antiproton Ring at CERN. He then joined the Laboratory and the LSND experiment as a postdoctoral staff member in 1995. He currently lives in Santa Fe and, in his spare time, enjoys mountain biking and working on his house.

Geoffrey Mills was born in Ithaca, New York, and grew up in Madison, Wisconsin. He received his B.S. in physics and mathematics from the University of Colorado. In 1985, Mills received his Ph.D. in experimental physics from the California Institute of Technology for research on the decay properties of the tau lepton. He then spent five years at the European Center for Nuclear Research studying the Z^0 boson at the Large Electron Positron collider. Since that time, Mills has been at the Laboratory pursuing the study of elementary particles while also skiing, hiking, playing the piano, and raising two boys with his wife, Ellen.

The Nature of Neutrinos in Muon Decay and Physics Beyond the Standard Model

Peter Herczeg

The main decay mode of the μ^+ [a] is the decay into a positron and two neutrinos: $\mu^+ \to e^+ + n + n'$ [b]. In the following we shall refer to these decays as "muon decay". Studies of muon decay played an important role in the developments that led to the $V - A$ theory of the weak interaction, and ultimately to the formulation of the electroweak component of the Standard Model (SM) [c]. Today the main motivation for further investigations of muon decay is to search for deviations from the predictions of the SM [1]. Although there is no definitive experimental evidence at present for physics beyond the SM, for many theoretical reasons, and especially because of the large number of undetermined parameters in the model, the existence of new physics is expected.

In the SM the interaction that mediates muon decay, as well as the nature of the neutrinos n and n', is prescribed. In extensions of the SM new interactions may contribute to the decay mode allowed in the SM, and there may be interactions that give rise to new decay modes of the type $\mu^+ \to e^+ + n + n'$. In the presence of new interactions the neutrinos are generally massive, and the weak eigenstates and the mass eigenstates of the neutrinos generally do not coincide.

Experiments, such as the KARMEN [2] and LSND [3] experiments, that search for $\bar{\nu}_e$'s originating from μ^+-decay are sensitive not only to oscillations of neutrinos into $\bar{\nu}_e$, but also to decays of the type $\mu^+ \to e^+ + \bar{\nu}_e + n_x$, [d] where n_x is a neutrino. The decays $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ are forbidden in the SM (they violate the conservation of lepton family numbers, and some of them also the conservation of the total lepton number). In this article we shall review these decays in some extensions of the SM. A question of importance is at what level of sensitivity searches for $\bar{\nu}_e$-appearance start to provide new information on the $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ branching ratios. The results of the LSND experiment brought added interest in this question. Could some of these branching ratios be large enough to account for the observed excess of e^+-events?

In any model, the decays $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ (and other two-neutrino muon decay modes) are constrained by muon-decay data obtained without observing the neutrinos and some also by experimental information on inverse muon decay. In the section below we describe these constraints, and also the experimental information on $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ from searches for $\bar{\nu}_e$-appearance. In the subsequent section we discuss the decays $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ in two important extensions of the SM: in a class of left-right symmetric models and in the minimal supersymmetric standard model with R-parity violation. The last section is a summary of the main points in the article.

---

a) We discuss for definiteness μ^+-decay. The general conclusions for μ^--decay are the same.

b) Unless otherwise stated, we use here the term "neutrino" for both neutrinos and antineutrinos. Thus, both n and n' can be a neutrino or an antineutrino (or a Majorana neutrino). In general μ^+ can have decay modes into several different pairs (n, n'). If n and/or n' are not mass eigenstates, then the decay for a given n and n' is into final states $e^+ + n_i + n_j'$, where n_i and n_j' are the various mass eigenstates contained in n and n'.

c) The electroweak component of the SM will be understood here to be the minimal version of the $SU(2)_L \times U(1)$ gauge theory, containing three families of leptons and quarks, one Higgs doublet, and only left-handed neutrinos.

d) The decays $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ are the only nonstandard types that can be studied experimentally at the available facilities, since the muons decay predominately at rest, and therefore the neutrinos are not energetic enough to produce μ's and τ's.

Muon Decay in the Standard Model: Comparison with Experiment

In the SM the neutrinos n and n' in muon decay are $n = \nu_{eL}$ and $n = \bar{\nu}_{\mu L}$, where $\nu_{lL} = (1 - \gamma_5)\nu_l/2$ $(l = e,\mu)$ are massless left-handed neutrinos (the ν_l are massless four-component spinors), which accompany the e^+ and the μ^+ in the decays $W^+ \to e^+ + \nu_{eL}$ and $W^+ \to \mu^+ + \nu_{\mu L}$. The latter decays result from the interaction

$$\mathcal{L} = \frac{g}{2\sqrt{2}} \left(\bar{\nu}_e \gamma_\lambda (1 - \gamma_5)e + \bar{\nu}_\mu \gamma_\lambda (1 - \gamma_5)\mu \right) W^\lambda + \text{H.c.,} \tag{1}$$

where the constant g is the gauge coupling constant associated with the $SU(2)_L$ factor of the SM gauge group. The structure of the interaction (1) is $V - A$, that is, the currents involved are given by the difference of the vector current $\bar{\nu}_l \gamma_\lambda l$ and the axial-vector current $\bar{\nu}_l \gamma_\lambda \gamma_5 l$. The interaction (1) generates the decay $\mu^+ \to e^+ + \nu_{eL} + \bar{\nu}_{\mu L}$ through the exchange of a W, as shown in Figure 1. The effective interaction describing the process in Figure 1 is given by

$$H_{SM} = \frac{g^2}{8m_W^2} \left(\bar{\mu}\gamma_\lambda (1 - \gamma_5)\nu_\mu \right)\left(\bar{\nu}_e \gamma_\lambda (1 - \gamma_5)e \right) + \text{H.c.,} \tag{2}$$

where m_W is the mass of the W ($m_W = 80.33 \pm 0.15$ GeV). What is the evidence regarding the description of muon decay in terms of the interaction (2)?

The presence of the W^- exchange contribution is certain: the W^- boson has been seen and studied, and its decays into $e^+ + \nu_e$ and $\mu^+ + \nu_\mu$ have been detected. A question of interest, which we shall consider now, is the fraction of the total muon-decay rate that can be attributed to this contribution.

The total muon-decay rate (the inverse of the muon lifetime) is given by

$$\Gamma^{(\mu)} \simeq \frac{G_F^2 m_\mu^5}{192\pi^3}. \tag{3}$$

In Equation (3) G_F is the Fermi constant. Its value is $G_F = 1.16639(2) \times 10^{-5}$ GeV^{-2} [4] deduced from measurements of the muon lifetime. In the absence of new contributions to muon decay, G_F is given by $G_F \simeq \sqrt{2}g^2/8m_W^2$. The constant $\sqrt{2}g^2/8m_W^2$ can be calculated using the experimental values of g and m_W. One obtains

$$\left(\frac{\sqrt{2}\,g^2}{8m_W^2} \right)^2 = (0.99 \pm 0.04)G_F^2. \tag{4}$$

Equation (4) implies that the observed muon-decay rate can be accounted for by the SM contribution. It also indicates that the SM contribution is the dominant one, unless there is a cancellation in the rate between the SM contribution and some new contributions.

Further information on the interactions involved in muon decay can be obtained from measurements on the positron (energy spectrum, polarizations, and angular distributions) and from the inverse muon decay processes $n_\pi + e^- \to \mu^- + n_a$, where n_a is a neutrino, and n_π is the neutrino emitted in the dominant $\pi^+ \to \mu^+ + neutrino$ decay.

If one assumes conservation of lepton family numbers, the only allowed two-neutrino decay mode of the μ^+ is $\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu$, where ν_e and ν_μ are in general Dirac neutrinos. The decay of the π^+ must proceed in this case as $\pi^+ \to \mu^+ + \nu_\mu$, and inverse muon decay as $\nu_\mu + e^- \to \mu^- + \nu_e$. In such a framework there are ten possible independent four-fermion interactions that can contribute to $\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu$. One of these is of the form (2) with $g^2/8m_W^2$ replaced by a general coupling constant $G_{LL}^V/\sqrt{2}$, where by the subscripts we have
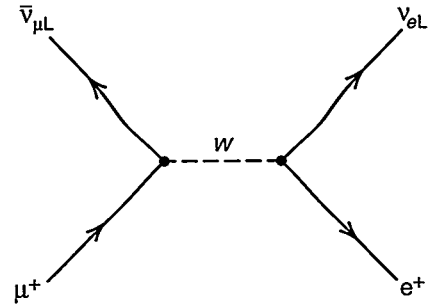

Figure 1.

indicated that the neutrinos in both currents are left-handed. One has $G_{LL}^V/\sqrt{2} = (g^2/8m_W^2) + \ldots$, where the dots stand for the coupling constants of other possible $V - A$ interactions. The Fermi constant is given now by $G_F^2 = |G_{LL}^V|^2 + |x|^2$, where $|x|^2$ represents the contributions of interactions with structures other than $V - A$. In this framework one of the conclusions of a combined analysis [5] of muon-decay data (with the neutrinos unobserved) and the inverse muon-decay cross section is that

$$|G_{LL}^V|^2 > 0.925\, G_F^2 . \tag{5}$$

This means that not more than about 10% of the muon-decay rate can originate from interactions other than $V - A$. In addition, upper bounds have been set on the coupling constants of other possible muon-decay interactions. The best of these is $\sim 3 \times 10^{-2} G_F$ [4].

The analysis mentioned above is not sufficiently general, since the assumption of lepton family number conservation is not justified: there is no reason why possible new contributions to muon decay should respect these conservation laws. A study of muon decay and inverse muon decay in the framework of a general interaction that allows for lepton family number violation and total lepton number violation has been carried out in Reference [6]. The conclusion is that the bounds obtained in the lepton family number conserving case for $|G_{LL}^V|^2$, $|G_{LL}^S|^2$,... apply in the general case for sums of the squares of certain combinations of the coupling constants. The sums which replace $|G_{LL}^V|^2$, $|G_{LL}^S|^2$, etc., contain G_{LL}^V, G_{LL}^S, etc., respectively.

The results of the general analysis imply [6] that at least one of the μ^+-decay modes which involve the neutrino state $\bar{n}_\pi$ (the neutrino state produced in the dominant $\pi^- \to \mu^- + neutrino$ decay) dominates the rate. We note yet that there is some experimental evidence (from pion decays) that n_π is not the neutrino state which accompanies the positron or the electron in nuclear beta decay. Some experimental information is available also on the second neutrino in muon decay. This comes from an experiment at LAMPF [7], in which the neutrinos from muon decay were detected for the first time. The detector when filled with heavy water was sensitive to neutrinos n_e capable of producing electrons in the reaction $n_e + D \to p + p + e^-$. The good agreement of the measured $n_e + D \to p + p + e^-$ cross section and the calculated one in the SM indicates that the total muon decay rate contains a substantial contribution from muon decay into a final state, in which one of the neutrinos is the one accompanying the positron in nuclear beta decay.

The evidence described above shows that the predictions of the SM for muon decay are consistent with experiment. Nevertheless, the data still leave room for relatively large (of the order of 10 percent in the rate) contributions from new physics.

Among the possible new μ^+-decay modes, the class characterized by the presence of $\bar{\nu}_e$ among the decay products can be identified by detecting the $\bar{\nu}_e$'s through the inverse beta-decay reaction $\bar{\nu}_e + p \to n + e^+$. Such experiments search for both $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ decays (where n_x is a neutrino), and for $\bar{\nu}_e$-appearance due to neutrino oscillations. A search for $\bar{\nu}_e$ originating from muon decay was carried out already in the experiment of Reference [7], where the detector was filled alternately with H_2O and D_2O. The experiment set an upper limit $B_{(\bar{\nu}_e)} < 0.098$ (90% c.l.) on the sum of the $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ branching ratios

$$B_{(\bar{\nu}_e)} \equiv \sum_{n_x} B(\mu^+ \to e^+ + \bar{\nu}_e + n_x) = \sum_{n_x} \Gamma(\mu^+ \to e^+ + \bar{\nu}_e + n_x)/\Gamma(\mu^+ \to all). \tag{6}$$

This limit was gradually improved by subsequent experiments [8]. The best present limit is [9]

$$B_{(\bar{\nu}_e)} < 2.5 \times 10^{-3} \qquad (90\% \text{ c.l.}), \qquad (7)$$

obtained in an experiment at the ISIS facility (Rutherford-Appleton Laboratory, United Kingdom) by the KARMEN collaboration. It should be noted that the limit (7), as well as all the previous limits, was derived under the assumption that the energy spectrum of the $\bar{\nu}_e$'s is the same as the energy spectrum of the $\bar{\nu}_\mu$'s in the SM muon decay. Some other possibilities are under study.

The probability $P_{\bar{\nu}_e}$ of $\bar{\nu}_e$ appearance found in the LSND experiment [3] is

$$P_{\bar{\nu}_e} = (0.31 \pm 0.13) \times 10^{-2}. \qquad (8)$$

If the excess of events found in the experiment is interpreted as due to $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ decays, $P_{\bar{\nu}_e}$ is the branching ratio $B_{\bar{\nu}_e}$ [Equation (6)]. It follows that

$$10^{-3} < B_{(\bar{\nu}_e)} < 5 \times 10^{-3} \quad (90\% \text{ c.l.}). \qquad (9)$$

This range is not inconsistent with the upper limit (7). The interpretations of the excess e^+-events in the LSND experiment in terms of $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ decays and in terms of neutrino oscillations are distinguishable since, unlike the branching ratio (6), the oscillation probability depends on the distance between the neutrino source and the detector and also on the neutrino energy.

The Decays $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ in Extensions of the Standard Model

The decays $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ occur in many extensions of the SM [10–13]. They can be mediated at the tree level by new gauge bosons, nonstandard Higgs bosons, and by the supersymmetric partners of the leptons. Here we shall consider these decays in a class of left-right symmetric models and in the minimal supersymmetric standard model with R-parity violation.

Left-Right Symmetric Models. These models [14] are attractive extensions of the electroweak sector of the SM. They provide a framework for the understanding of parity violation in the weak interaction. In the SM parity violation is introduced in an ad hoc manner, by arranging (following experiment) that the W-boson couples only to currents involving the left-handed components of the fermion fields. These couplings have a form analogous to those in Equation (1). The question of why nature appears to select fermions of only one handedness to participate in the weak interaction is in the context of the SM unanswered. In left-right symmetric models (LRSM) parity violation appears in a new light. The gauge group of the simplest LRSM is $SU(2)_L \times SU(2)_R \times U(1)$, which is larger than the gauge group of the SM by the $SU(2)_R$ factor. The observed W-boson (called here W_L) is associated with $SU(2)_L$, while the $SU(2)_R$ group accommodates a second (hitherto undetected) charged gauge boson, the W_R, which couples only to right-handed currents (i.e., currents involving the right-handed components of the fermion fields). The model requires the existence of right-handed neutrinos ν'_{lR} ($l = e,\mu,\tau$), which are the partners of the right-handed components of the charged leptons l_R^- in doublets of $SU(2)_R$. Thus, the right-handed neutrinos are not sterile, but participate in the right-handed interactions. Also, the neutrinos are expected to have nonzero masses.

The general effect of the W_R can be illustrated on the example of muon decay. The exchange of the W_R gives a second contribution to $\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu$ (see Figure 2), [e] which is of the $V + A$ form
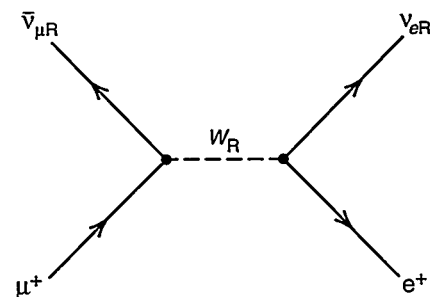

Figure 2.

---

e) For simplicity we are taking for this argument the neutrinos to be Dirac particles, and neglecting neutrino mixing.

$$H_{W_R} = \frac{g_R^2}{8m_{W_R}^2}\left(\bar{\mu}\gamma_\lambda\left(1 + \gamma_5\right)\nu_\mu\right)\left(\bar{\nu}_e\gamma_\lambda\left(1 - \gamma_5\right)e\right) + \text{H.c. .} \tag{10}$$

In Equation (10) g_R is the coupling constant associated with $SU(2)_R$; the coupling constant associated with $SU(2)_L$ is denoted now by g_L; m_{W_R} is the mass of W_R. If $g_R = g_L$, then in the limit when the masses of the W_L and W_R are equal, the sum of the effective interactions (2) and (10) conserves parity. The observed parity violation arises through spontaneous symmetry breaking (the same mechanism which generates the mass of the W in the SM). This can make the W_R heavier than the W_L. The left-handed interactions then dominate, but parity violation is no longer maximal, since the right-handed interactions also participate. The strength of the right-handed interactions depends on the size of m_{W_R}. The present experimental lower bound on m_{W_R} is at least 300 GeV, and in most versions of $SU(2)_L \times SU(2)_R \times U(1)$ models larger. This means that the right-handed interactions must be weaker than the usual weak interactions by at least an order of magnitude.
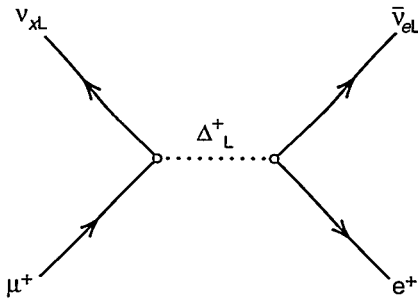
The Higgs sector of $SU(2)_L \times SU(2)_R \times U(1)$ models is richer than that of the SM, in part because the gauge symmetry, which has to be broken to electromagnetic gauge invariance only, is larger. An attractive choice for implementing the symmetry breaking is a Higgs sector that includes two triplets of Higgs bosons, $\vec{\Delta}_R \equiv \left(\Delta_R^{++}, \Delta_R^+, \Delta_R^0\right)$ and $\vec{\Delta}_L \equiv \left(\Delta_L^{++}, \Delta_L^+, \Delta_L^0\right)$, which couple to the right-handed and the left-handed leptons, respectively. With this choice, and with some additional assumptions, the model predicts a seesaw relation of the form, $m_{\nu_l} \propto m_l^2/m_{W_R}$ for the neutrino masses. In this version of LRSM the right-handed neutrinos are heavy, with masses of the order of m_{W_R}.

In the above version of $SU(2)_L \times SU(2)_R \times U(1)$ models in addition to the usual muon-decay mode $\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu$ the lepton family number violating decays $\mu^+ \to e^+ + \bar{\nu}_e + \nu_x$ $(x = e,\mu,\tau)$ also occur [11]. They are mediated by the singly charged component of the $\vec{\Delta}_L$, as shown in Figure 3.

We shall assume in the following that mixing in the leptonic sector can be neglected. Then the decay $\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu$ [f] will be the dominant one, since it is the only one which survives in the absence of family mixing. The effective interaction responsible for this decay is given by

$v_{\mu L}$ $\bar{v}_{eL}$

Δ^+_L

μ^+ e^+

Figure 3.

$$H_\Delta = \frac{G'}{\sqrt{2}}\left(\bar{\mu}\left(1 + \gamma_5\right)\nu_\mu^c\right)\left(\bar{\nu}_e^c\left(1 - \gamma_5\right)e\right) + \text{H.c.,} \tag{11}$$

where $G' = -\sqrt{2}f_{ee}\,f_{\mu\mu}^*/2m_+^2$, f_{ee} and $f_{\mu\mu}$ are $\vec{\Delta}_L$-lepton coupling constants, and m_+ is the mass of the Δ_L^+; the field ν_l^c $(l = e,\mu)$ describes the right-handed antiparticle of ν_l. The branching ratio $B(\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu) \approx B_{(\bar{\nu}_e)}$ is given by

$$B(\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu) \approx \frac{1}{4}\left|\frac{G'}{G_F}\right|^2. \tag{12}$$

From the experimental limit (7) one obtains

---

[f] The decay $\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu$ was first considered [15] before the advent of gauge theories, in connection with the question regarding the nature of the suspected conservation law, which was supposed to account for the apparent absence of processes like $\mu \to e\gamma$. The decay $\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu$ would be allowed if the absence of $\mu \to e\gamma$ is due to the conservation of a particular multiplicative quantum number ("muon parity"), while both $\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu$ and $\mu \to e\gamma$ are forbidden by the conservation of the additive "muon number" (which is identical to the muon family number). We note that since an interaction responsible for $\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu$ is not the weak interaction, the existence of a conserved multiplicative quantum number cannot be ruled out by the absence of $\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu$ at a certain level.

$$|G'| < 10^{-1} G_F. \tag{13}$$

A bound on G' follows also from a new experimental limit on muonium to antimuonium conversion [16]. Muonium (antimuonium) is a bound state of $\mu+$ and e^- $(\mu^-$ and $e^+)$. In the model we are considering, muonium to antimuonium conversion is mediated by the doubly charged component of $\vec{\Delta}_L$ [17]. The coupling constant $G_{M\overline{M}}$ of the corresponding interaction is related to G' as $G' \simeq -4G_{M\overline{M}}\, m_{++}^2/m_+^2$, where m_{++} is the mass of the Δ_L^{++}. Assuming that the mixing of $\vec{\Delta}_L$ with other Higgs fields can be neglected, one has the relation $m_+^2 = (m_{++}^2 + m_0^2)/2$ [18] among the masses of the Δ_L^{++}, Δ_L^+, and Δ_L^0. This relation and the experimental limit $|G_{M\overline{M}}| < 3 \times 10^{-3}$ (90% c.l.) [16] imply

$$|G'| < 2.4 \times 10^{-2} G_F, \tag{14}$$

and therefore

$$B(\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu) < 1.5 \times 10^{-4}, \tag{15}$$

which is too small to account for the LSND result. The consequences of mixing among the Higgs fields and in the leptonic sector are under study.

In $SU(2)_L \times SU(2)_R \times U(1)$ models the decay $\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu$, and muonium to antimuonium conversion, can give important information on the values of the ν_μ-mass allowed in these models [11]. As in the SM, the neutrino masses in $SU(2)_L \times SU(2)_R \times U(1)$ models are undetermined. In any model, the masses and lifetimes of the neutrinos are constrained by the requirement that in the present universe the energy density of the neutrinos does not exceed the upper limit on the total energy density of the universe. This can be shown to imply that neutrinos of masses between ~35 eV and ~3 GeV have to be unstable. For such neutrinos there is a relation between their masses and lifetimes. The heavier the neutrino, the faster it has to decay. An issue of interest is then whether in a given model there is a decay mode which allows a given neutrino to decay fast enough. For ν_μ in $SU(2)_L \times SU(2)_R \times U(1)$ models the only such decay mode turns out to be $\nu_\mu \to \nu_e + \nu_e + \bar{\nu}_e$, and this only for ν_μ's with masses in the range

$$40 \text{ keV} \lesssim m_{\nu_\mu} < 170 \text{ keV}. \tag{16}$$

The upper limit in Equation (16) is the present experimental limit on m_{ν_μ}. Can the ν_μ have a mass in the range of (16)? The special role of the decay $\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu$ and of muonium to antimuonium conversion in $SU(2)_L \times SU(2)_R \times U(1)$ models is that they can probe this question. The dominant mechanism for the decay $\nu_\mu \to \nu_e + \nu_e + \bar{\nu}_e$ is the exchange of the neutral component of the $\vec{\Delta}_L$. For the decay rate to be sufficiently large, Δ_L^0 cannot be arbitrarily heavy. As follows from some further considerations, this implies that the constant G' has a lower bound for m_{ν_μ}'s in the range (16). This lower bound is $|G'| \gtrsim 7 \times 10^{-4}$, to be compared with the bound (14). It can be shown that as the experimental limit on the $B(\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu)$ becomes more and more stringent than the bound (15), the lower bound for the possible values of m_{ν_μ} in Equation (16) will become increasingly larger.

The Minimal Supersymmetric Standard Model with R-parity Violation. Supersymmetry is an extension of the known space-time symmetries (the invariance with respect to the inhomogeneous Lorentz transformations) [19]. Supersymmetry transforms bosons (fermions) into fermions (bosons). The supersymmetric version of the SM, the minimal supersymmetric standard model (MSSM) contains not only the SM fields, but also their superpartners. The superpartners of the leptons and

the quarks, for example, are spin-zero particles, called the sleptons and the squarks, respectively. Supersymmetric models can solve some of the theoretical problems of the SM and allow the unification of the interactions of the SM with gravity.

Unlike in the SM, in the MSSM the conservation of lepton (L) and baryon (B) numbers is not automatic: the Lagrangian can contain L- and B-violating gauge-invariant supersymmetric terms. To eliminate B-violation, which would have to be extremely weak to prevent too rapid proton decay, a discrete symmetry, called R-parity symmetry, is usually imposed [20]. R-parity is a multiplicative quantum number, whose value is $+1$ for the SM particles, and -1 for the super-partners. The requirement of R-parity invariance eliminates not only the B-violating terms, but also the L-violating ones. Alternatively, with a different choice of the discrete symmetry, it is possible to arrange that the L-violating terms remain. The presence of R-parity violating terms in the Lagrangian has rich phenomeno-logical consequences. One of these is that they allow the production of single superpartners (in R-parity-conserving models the superpartners have to be pro-duced in pairs). Another is that they give rise to some new processes that are for-bidden in the SM. Among these are new two-neutrino decays of the muon.

The decay $\mu^+ \rightarrow e^+ + \bar{\nu}_e + \nu_\mu$ has been considered in this model in Reference [12]. It is mediated by the superpartner $\tilde{\tau}_L$ of the left-handed component of the τ, as shown in Figure 4. The corresponding interaction is of the form
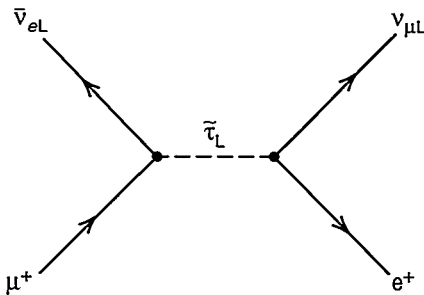
$$H_{\tilde{\tau}} = \frac{G''}{\sqrt{2}}\left(\bar{\mu}(1 - \gamma_5)\nu_e\right)\left(\bar{\nu}_\mu(1 + \gamma_5)e\right) + \text{H.c.} , \qquad (17)$$

Figure 4.

where $G''/\sqrt{2} = \lambda_{132}\,\lambda^*_{231}/4m_{\tilde{\tau}}^2$, $m_{\tilde{\tau}}$ is the mass of the $\tilde{\tau}_L$, and the λ's are cou-pling constants; the subscripts on the λ's are family indices. The product of the λ's in Equation (17) turns out to govern also muonium to antimuonium conversion [12]. The latter process is mediated by $\tilde{\nu}_\tau$ the superpartner of the ν_τ. From the experimental limit [16] on muonium to antimuonium conversion we obtain the bound g)

$$B\left(\mu^+ \rightarrow e^+ + \bar{\nu}_e + \nu_\mu\right) \lesssim 10^{-4} , \qquad (18)$$

which is much below the LSND range (9).

In addition to $\mu^+ \rightarrow e^+ + \bar{\nu}_e + \nu_\mu$, other decays of the type $\mu^+ \rightarrow e^+ + \bar{\nu}_e + n_x$ also occur. An analysis [21] shows that under the assumption that lepton and slepton mixing is small, the sum of their branching ratios is $B_{(\bar{\nu}_e)} \lesssim 2 \times 10^{-4}$. The effects of mixings are being investigated.

Summary

In this article we discussed two-neutrino decays of the muon of the type $\mu^+ \rightarrow e^+ + \bar{\nu}_e + n_x$, where n_x is a neutrino. Such decays are experimentally accessible through the detection of $\bar{\nu}_e$'s using the inverse beta-decay reaction $\bar{\nu}_e + p \rightarrow e^+ + n$. The decays $\mu^+ \rightarrow e^+ + \bar{\nu}_e + n_x$ are of considerable importance, since they probe leptonic interactions that are not present in the Standard Model. A new issue of interest is whether such decays could be fast enough to be potential sources of the observed excess of e^+ events in the LSND experiment. In this connection it should be noted that the data from a recent experiment [22] of the LSND collaboration on

g) We used $m_{\tilde{\tau}}^2 - m_{\nu_{\tilde{\tau}}}^2 \lesssim 0.8\, m_{\tilde{\tau}}^2$ (see P. Nath et al., Reference [19]) and $m_{\tilde{\tau}} \approx 45$ GeV (the experi-mental lower bound on $m_{\tilde{\tau}}$ see Reference [4]).

$\nu_\mu \to \nu_e$ oscillations support the interpretation of the result of their previous experiment in terms of $\bar{\nu}_\mu \to \bar{\nu}_e$ oscillations.

We reviewed the decays $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ in two important extensions of the Standard Model: in a class of left-right symmetric models and in the minimal supersymmetric standard model with R-parity violation. In the version of left-right symmetric models where the smallness of the masses of the usual neutrinos is related to the large size of the scale of the right-handed interactions, the decay $\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu$ (which is expected to dominate among $\mu^+ \to e^+ + \bar{\nu}_e + n_x$) is mediated by the singly charged component of a triplet of Higgs bosons $\vec{\Delta}_L$. Assuming that the mixing of $\vec{\Delta}_L$ with other Higgs fields can be neglected, the upper limit on the $\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu$ branching ratio turns out to be $\sim 10^{-4}$ (implied by the present limit on muonium to antimuonium conversion). This is an order of magnitude below the range required to account for the LSND result. In the minimal supersymmetric standard model with R-parity violation the upper limit on the sum of the $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ branching ratios is $\sim 2 \times 10^{-4}$, assuming that lepton and slepton mixing is small. A further study of $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ decays in the above models, and also in other extensions of the standard model is in progress.

To improve the sensitivity of experiments searching for $\mu^+ \to e^+ + \bar{\nu}_e + n_x$ decays remains important. In the case of left-right symmetric models improved limits on the $\mu^+ \to e^+ + \bar{\nu}_e + \nu_\mu$ branching ratio would also provide information on the mass of the muon neutrino in these models. ∎


Peter Herczeg received his M.S. from the Czech Technical University in Prague, Czech Republic, and his Ph.D. from the University of Sussex, England. During 1962–1965 he was a staff member at the Institute of Physics of the Czechoslovak Academy of Sciences in Prague. Subsequently he held research associate positions at the Niels Bohr Institute (Copenhagen), University of Sussex, California Institute of Technology, and at Carnegie-Mellon University. In 1973 he joined the Medium Energy Physics Group (T-5) of the Theoretical Division at Los Alamos. His research has been mainly in the areas of the weak interactions and the phenomenology of extensions of the Standard Model of the fundamental interactions. It has included subjects in the fields of CP violation, rare decays, and time reversal violation in low-energy processes. For his contributions to these fields he was elected in 1993 a Fellow of the American Physical Society.

References

[1] For reviews of muon decay see W. Fetscher and H.-J. Gerber, in *Precision Tests of the Standard Electroweak Model*, Advanced Series on Directions in High Energy Physics, edited by P. Langacker (World Scientific, Singapore, 1995), p. 657; P. Herczeg, ibid., p. 786, and Z. Phys. C *56*, S129 (1992).

[2] See K. Eitel (for the KARMEN Collaboration), *hep-ex/9706023* (June 1997).

[3] C. Athanassopoulos, et al. (LSND Collaboration), Phys. Rev. C *54*, 2685 (1996); Phys. Rev. Lett. *75*, 2650 (1995). See also W. Louis, V. Sandberg, and H. White, "A Thousand Eyes—The Story of LSND" on page 92 in this issue.

[4] Particle Data Group, R. M. Barnett et al., Phys. Rev. D *54*, (July 1996), and the 1997 update.

[5] W. Fetscher, H.-J. Gerber, and K. F. Johnson, Phys. Lett. B *173*, 102 (1986); W. Fetscher and H.-J. Gerber, Reference [1], p. 251.

[6] P. Langacker and D. London, Phys. Rev. D *39*, 266 (1989).

[7] S. E. Willis et al., Phys. Rev. Lett. *44*, 522 (1980), *45*, 1370 (E) (1980).

[8] D. A. Krakauer et al., Phys. Lett. B *263*, 534 (1991); S. Freedman et al., Phys. Rev. D *47*, 811 (1993); K. Eitel, Ph.D. thesis, Universität und Forschungszentrum Karlsruhe (1995), Forschungszentrum Karlsruhe Report FZKA 5684 (1995).

[9] K. Eitel, Reference [8].

[10] P. H. Frampton and D. Ng, Phys. Rev. D *45*, 4240 (1992); H. Fujii, S. Nakamura, and K. Sasaki, Phys. Lett. B *299*, 342 (1993); H. Fujii, Y. Mimura, K. Sasaki, and T. Sasaki, Phys. Rev. D *49*, 559 (1994).

[11] P. Herczeg and R. N. Mohapatra, Phys. Rev. Lett. *69*, 2475 (1992).

[12] A. Halprin and A. Masiero, Phys. Rev. D *48*, R2987 (1993).

[13] G.-G. Wong and W-S. Hou, Phys. Rev. D *50*, R2962 (1994).

[14] J. C. Pati and A. Salam, Phys. Rev. D *10*, 275 (1974); R. N. Mohapatra and J. C. Pati, Phys. Rev. D *11*, 566 (1975); G. Senjanović and R. N. Mohapatra, Phys. Rev. D *12*, 1502 (1975); R. N. Mohapatra and G. Senjanović, Phys. Rev. Lett. *44*, 912 (1980) and Phys. Rev. D *23*, 165 (1981).

[15] G. Feinberg and S. Weinberg, Phys. Rev. Lett. *6*, 381 (1961).

[16] V. Meyer (for the Muonium—Antimuonium Collaboration), paper presented at the *Sixth Conference on the Intersections of Particle and Nuclear Physics (CIPANP97)*, Big Sky, Montana, May 27–June 2, 1997.

[17] A. Halprin, Phys. Rev. Lett. *48*, 1313 (1982).

[18] R. N. Mohapatra and P. B. Pal, Phys. Lett. *179* B, 105 (1986); R. N. Mohapatra and G. Senjanović, Phys. Rev. D *23*, 165 (1981); C. S. Lim and T. Inami, Prog. Theor. Phys. *67*, 1569 (1982); N. G. Deshpande, J. F. Gunion, B. Kayser, and F. Olness, Phys. Rev. D *44*, 837 (1991).

[19] For reviews of supersymmetry see, for example, H. E. Haber in *Recent Directions in Particle Theory*, Proceedings of the 1992 Theoretical Advanced Study Institute in Particle Physics, edited by J. Harvey and J. Polchinski (World Scientific, Singapore, 1993), p. 589; R. N. Mohapatra, *Unification and Supersymmetry* (Springer Verlag, New York, 1986); P. Nath, R. Arnovitt, and A. H. Chamseddine, *Applied N = 1 Supergravity* (World Scientific, Singapore, 1984); H. E. Haber and G. L. Kane, Phys. Reports *117*, 75 (1985); H. P. Nilles, Phys. Reports *110*, 1 (1984).

[20] Recent reviews of supersymmetric models with R-parity violation include H. Dreiner, *hep-ph/9707435* (July 1997); G. Bhattacharya invited talk at the workshop *Beyond the Desert — Accelerator and Non-Accelerator Approaches*, Castle Ringberg, Tegernsee, Germany, June 8–14, 1997, to appear in the proceedings to be published by World Scientific.

[21] P. Herczeg, "Exotic Muon Decays and the LSND Result on $\bar{\nu}_\mu \to \bar{\nu}_e$ Oscillations", in preparation.

[22] C. Athanassopoulos et al. (LSND Collaboration), "Evidence for $\nu_\mu \to \nu_e$ Oscillations from Pion Decay in Flight Neutrinos", *nucl-ex/9706006*, submitted to Phys. Rev. C. See also W. Louis et al., Reference [3].
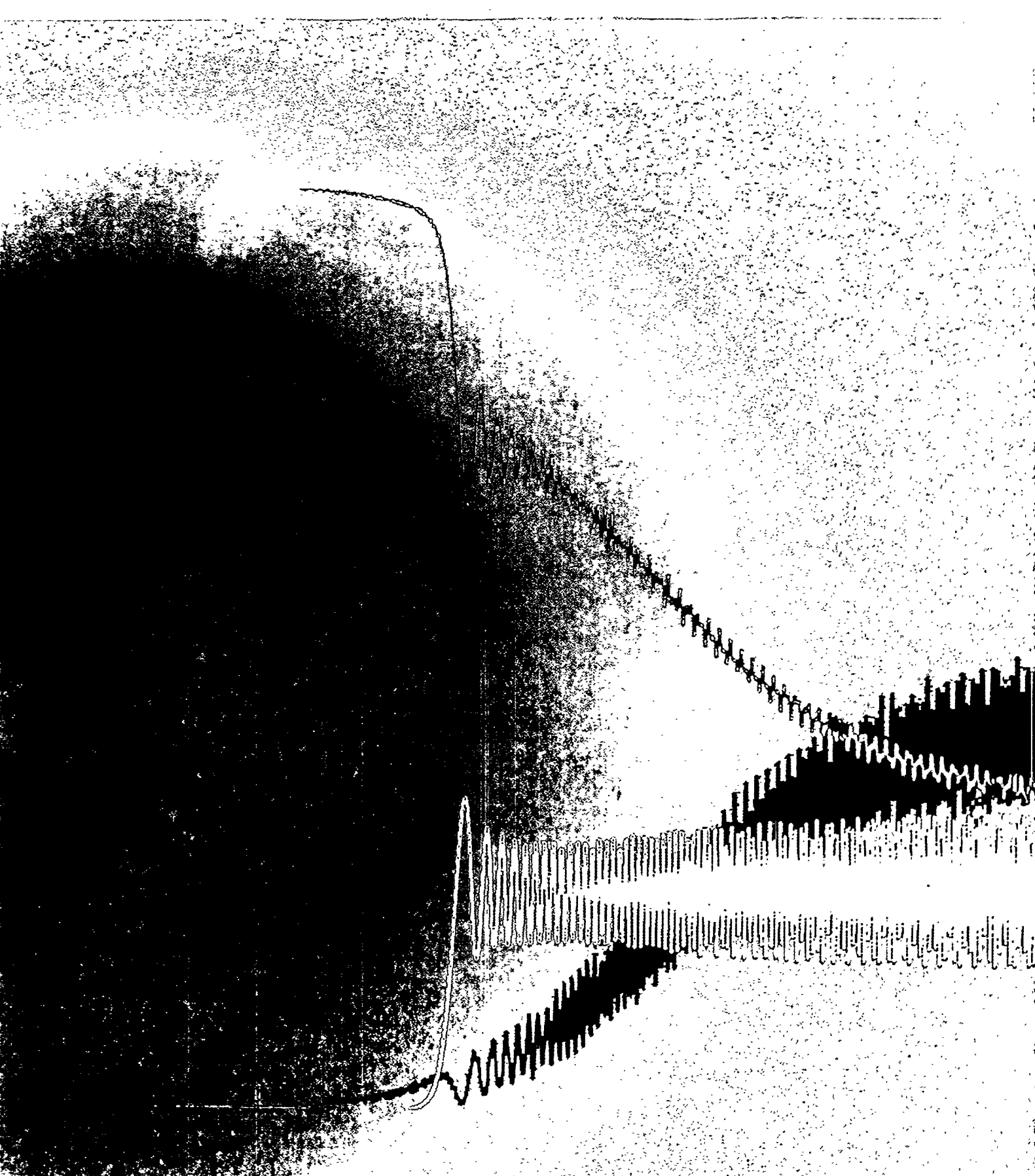
If neutrinos have mass, then the three separate particles known as the electron neutrino, the muon neutrino, and the tau neutrino may not be separate at all, but may mix and transform into one another. In this illustration, a large fraction of the electron neutrinos produced in the core of the sun change their identity before they reach the surface (blue curve). They reappear either as muon and/or tau neutrinos (red and yellow curves, respectively).

Exorcising Ghosts

In pursuit of the missing solar neutrinos

Andrew Hime

After thirty years of hints that electron neutrinos slip in and out of existence, new solar-neutrino experiments may finally catch them in the act.

In October 1920, Sir Arthur Eddington, one of the foremost astrophysicists of the century, delivered his presidential address to the British Association at Cardiff. In his speech, entitled "The Internal Constitution of the Stars," he referred to a proposal suggested the year before by the former president of the association to bore a hole into the crust of the earth in order to discover the conditions deep below. Motivated by the rapid progress in astronomy at the time, Sir Eddington proposed something "easier" to penetrate, namely, the Sun.

Eddington could scarcely have anticipated the ramifications of his suggestion. After more than seventy-five years of study, the scientific community's investigations of our closest star have yielded a remarkably detailed understanding of what makes the Sun shine. We now know that the Sun is powered by thermonuclear fusion and that its hot core can be considered an immense furnace producing not only heat and light, but also vast numbers of neutrinos.

Because of the Sun's enormous size, the light produced deep in its interior takes tens of years to reach its surface. During that lengthy journey, the photons that rain down upon us as sunlight and make our existence on Earth sustainable lose all the information concerning the detailed processes of the stellar core. Unlike photons, neutrinos interact so feebly with matter that they escape from the Sun in about 2 seconds. They arrive on Earth a mere 8 minutes later, and thus the solar neutrinos are a unique probe of a star's innermost regions and of the nuclear reactions that fuel them.

During the past thirty years, detailed theoretical and experimental studies have resulted in very precise predictions about the fluxes and energy spectra of neutrinos produced deep within the Sun. But a problem has emerged. Four different experiments have measured the flux of solar neutrinos, and every one of them reports a flux that is significantly below theoretical predictions. The discrepancy is referred to as the *solar-neutrino problem*, and it

is particularly puzzling because scientists have failed to find errors in the standard theoretical framework of the Sun or in the terrestrial experiments monitoring the neutrinos.

Where have the solar neutrinos gone? One intriguing answer may lie outside our conventional understanding of physics. Whereas a remedy based upon modifications in solar models appears difficult to construct, scientists are particularly excited about the possibility that something profound may happen to the neutrinos as they make their way out of the Sun en route to Earth.

We know of three different types, or flavors, of neutrinos—the electron, muon, and tau neutrinos. We also know that the nuclear reactions that power the Sun are energetic enough to produce only electron neutrinos. Moreover, existing experiments that detect solar neutrinos are only sensitive to the electron flavor. One can thus speculate that some of the electron neutrinos produced in the Sun have transformed, or oscillated, into muon and/or tau neutrinos as they make their way to Earth, thereby escaping our terrestrial detectors. The probability for oscillations to occur may even be enhanced in the Sun in an energy-dependent and resonant manner as neutrinos emerge from the dense core. This phenomenon, an example of the Mikheyev, Smirnov, and Wolfenstein (MSW) effect, is considered by many scientists to be the most favored solution to the solar-neutrino problem.

Neutrino oscillations, or the periodic changes in neutrino flavor, require that neutrinos possess mass and that neutrino flavor not be conserved in nature. No undebated evidence for neutrino mass exists despite years of painstaking research around the world. Indeed, the Standard Model of elementary particles requires that neutrinos be strictly massless. Nonetheless, quests for a Grand Unified Theory of the fundamental forces in nature suggest that neutrinos, like other elementary particles, should have mass. Consequently, should the solar-neutrino problem be resolved by invoking neutrino mass and oscillations,

the result would be evidence for physics beyond the Standard Model. The models that emerge from elementary particle physics, astrophysics, and cosmology would be subject to a new set of constraints and would have to be modified with potentially profound implications.

The status of the solar-neutrino problem, along with how new experiments propose to solve it, forms the central theme of this article. Particular emphasis is reserved for the Sudbury Neutrino Observatory, an experiment under construction that promises to resolve the question of whether neutrino oscillations, and in particular the MSW effect, are responsible for the observed shortfall of solar neutrinos.

Neutrinos from the Sun

Given the enormous power produced by the Sun and its twenty-billion-year lifetime, it is a steadfast conclusion that the Sun produces energy via thermonuclear fusion. During the late 1920s and early 1930s, theoretical calculations, including the seminal work of a young Hans Bethe, elucidated our understanding of the details of these processes. As shown in Figure 1, the fusion of protons into helium proceeds via three branches. Neutrinos are created in four different reactions, referred to simply as the *pp*, *pep*, beryllium-7 (^{7}Be), and boron-8 (^{8}B) reactions. The neutrinos flee the Sun and begin their voyage to Earth. (In Figure 1, we have omitted neutrinos that emerge from the carbon-nitrogen-oxygen, or CNO, cycle. The cycle is another, though less important, set of neutrino-producing reactions in the Sun.)

Figure 2 shows the predictions of the standard solar model for the flux of electron neutrinos at the earth's surface. The flux is the number of neutrinos per square centimeter per second. (The figure assumes no electron neutrinos have oscillated into a different flavor.) The *pp* reaction is the primary mode of neutrino production, and the reaction completely dominates energy production in the Sun.
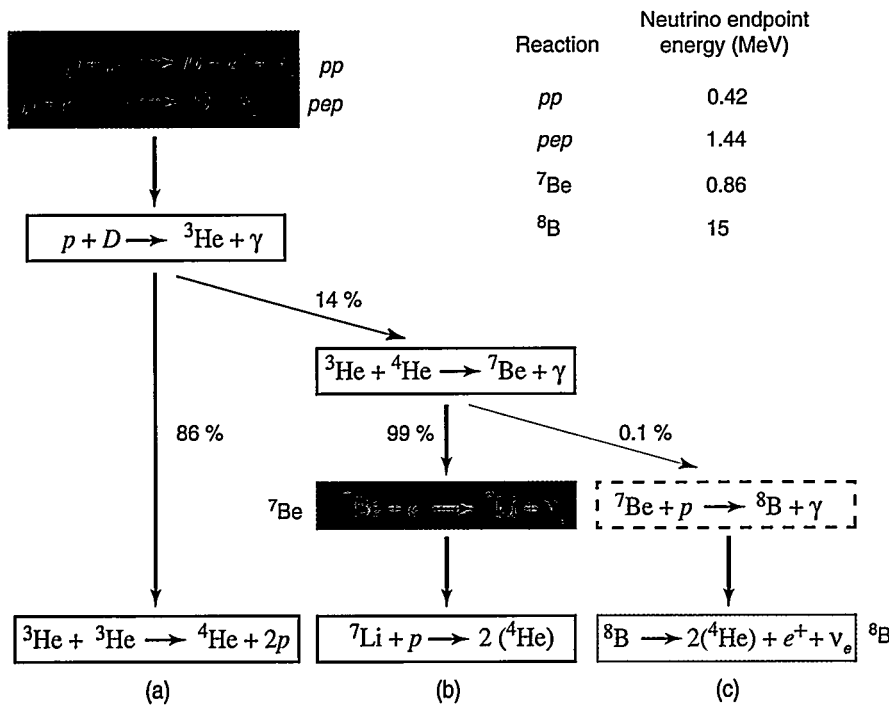
Reaction	Neutrino endpoint energy (MeV)
pp	0.42
pep	1.44
^7Be	0.86
^8B	15

pp

pep

$$p + D \longrightarrow {}^3He + \gamma$$

14 %

$${}^3He + {}^4He \longrightarrow {}^7Be + \gamma$$

86 %   99 %   0.1 %

^7Be $^7Be + p \longrightarrow {}^8B + \gamma$

$$^3He + {}^3He \longrightarrow {}^4He + 2p$$

$$^7Li + p \longrightarrow 2\,({}^4He)$$

$$^8B \longrightarrow 2({}^4He) + e^+ + \nu_e$$ ^8B

(a)   (b)   (c)

Figure 1. The Primary Neutrino-Producing Reactions in the Sun
Nearly all the Sun's energy comes from the fusion of protons into deuterium nuclei. The deuterium is converted into helium-4 by following one of three reaction pathways (labeled a, b, and c). Of the reactions shown, four proceed via charge-changing weak interactions (colored boxes) and therefore produce electron neutrinos. Over 95 percent of the neutrinos are created in the pp fusion reaction. One proton undergoes inverse beta decay, creating a neutron, positron, and an electron neutrino. The neutron then binds to the proton to form a deuteron (labeled D). Other neutrino-producing reactions are pep (electron capture), ^7Be (electron capture), and ^8B (beta decay). Notice that ^7Be is needed to produce ^8B (dashed box). Modern experiments, however, observe neutrinos from the pp reaction and ^8B decay, but hardly any from ^7Be decay.

Range of Sensitivity

0.23   0.8   7.0

Gallium

Chlorine

Kamiokande

10^{11}

$\leftarrow pp$

Neutrino Flux (number /cm^2/s)

10^9 CNO (^{13}N)

$\leftarrow {}^7Be$

10^7 CNO (^{15}O)

$\leftarrow pep$

10^5 CNO (^{17}F)

$\leftarrow {}^8B$

10^3

hep

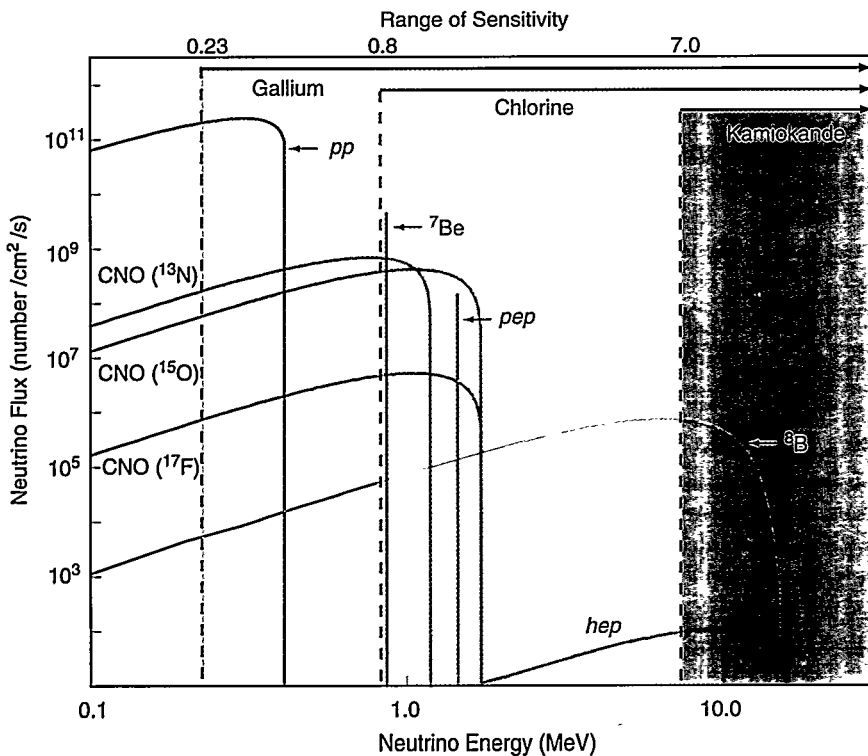0.1   1.0   10.0

Neutrino Energy (MeV)

Figure 2. Solar-Neutrino Spectrum
The total integrated flux of all solar neutrinos reaching the earth is about 65 billion per square centimeter per second. In this figure, the neutrino flux and energy are plotted on log scales; so, for example, the pp flux is about 50 times greater than the ^7Be flux. Also shown are the spectra of neutrinos produced from the CNO cycle (gray curves). The pp, ^8B, and CNO neutrinos are created in beta decay reactions. A neutrino so produced shares energy with another light particle. Hence, all those neutrinos have a broad energy spectrum. The ^7Be and pep neutrinos result from electron capture: A proton in a nucleus captures an electron from an atomic orbital, turns into a neutron, and a monoenergetic neutrino is created. The sensitivity range of the various solar-neutrino experiments is also shown here. The gallium experiments have energy thresholds around 0.23 MeV and are sensitive to all solar neutrinos. The chlorine experiment detects neutrinos from the ^7Be and CNO reactions, but is primarily sensitive to those from the ^8B reaction. Kamiokande is a water Cerenkov detector that can detect only the high-energy portion of ^8B neutrinos.

Consequently, it is ultimately linked to many solar observables. Indeed, measurements of the light output of the Sun (its luminosity) place rather stringent constraints on the rate of the *pp* reaction. As such, a prediction of the associated *pp* neutrino flux hardly changes as the details of a solar model change.

But neutrino fluxes from the ^{7}Be and ^{8}B reactions are many times smaller than the *pp* flux. The magnitude of those fluxes depends critically on the details of a solar model and in particular on the physical inputs that can alter the central temperature of the Sun (typically calculated to be about 15 million kelvins). Specifically, the ^{8}B flux, which is manifested as a mere one hundredth of 1 percent of the total flux of solar neutrinos, varies as the 25th power of the core temperature. Hence, a variation in this temperature of only a few percent can drastically alter the prediction of the ^{8}B neutrino flux.

Four pioneering experiments have measured the solar-neutrino flux. These experiments were designed to detect neutrinos via their interactions with nuclei or electrons in the target material making up the host detector. Because the target materials were different, the experiments were sensitive to different neutrino energies, as indicated in Figure 2. Taken together, the experiments that are discussed below have provided a coarse mapping of the entire solar-neutrino spectrum.

The Pioneering Experiments

Given data on the Sun's luminosity, the standard solar model provides a rather straightforward prediction for the total flux of solar neutrinos reaching the earth. Specifically, some 6.57×10^{10} are predicted to pass through every square centimeter of our planet each second. Despite this impressive number, the neutrinos interact so weakly with matter that the probability of detecting any one of them is miniscule. Another way to state the problem is that an atom presents an extremely tiny target

to the neutrino. Hence, to have any hope of catching a neutrino, one either waits a long time or builds a monstrous detector that contains a huge number of target atoms.

Solar-neutrino experiments exploit both strategies. The experiments run for years and make use of detectors that contain hundreds to thousands of tons of target material. Even so, neutrino interactions are still rare in these watchful behemoths. Typically, one expects to record only a few events *per day*! Consequently, the detectors are buried under mountains or burrowed into mine shafts in order to prevent cosmic rays from striking the target and inducing background signals. The locations underscore the ironic truth surrounding solar-neutrino experiments—one can move deep into the earth and still see the sun shine!

The Chlorine Experiment. This ground-breaking experiment has been in progress for nearly thirty years. Situated 4,500 feet underground in the Homestake Mine in South Dakota, the experiment uses a large tank of perchloroethylene (C_2Cl_4), a common dry-cleaning fluid, to snare the ghostly neutrinos.

Electron neutrinos ν_e from the Sun make themselves known through the inverse-beta-decay reaction on chlorine nuclei:

$$\nu_e + {}^{37}\text{Cl} \rightarrow {}^{37}\text{Ar} + e^- \ . \qquad (1)$$
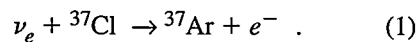
A neutron in the chlorine-37 (^{37}Cl) nucleus is transformed into a proton to yield the daughter argon-37 (^{37}Ar) nucleus with the emission of an electron (e^-). The neutrino must have at least 0.814 million electron volts (MeV) of energy in order to drive this reaction. This means (see Figure 2) that the chlorine experiment is sensitive to neutrinos from the ^{7}Be, *pep*, and ^{8}B reactions. Although the number of ^{7}Be neutrinos is predicted to be far greater than that of ^{8}B neutrinos, the largest signal in the chlorine experiment is due to the ^{8}B neutrinos. This is because the daughter ^{37}Ar nucleus has an excited

state that can be accessed only by the high-energy neutrinos present in the ^{8}B spectrum. Consequently, those neutrinos have a larger interaction-cross-section and are detected at a greater rate.

Once produced, the unstable ^{37}Ar atoms eventually decay by recapturing an orbital electron (half-life approximately 35 days) and become once again ^{37}Cl. Characteristic x-rays are emitted that signal the decay, but they can only be detected after the ^{37}Ar atoms have been extracted from the chlorine tank. Bubbling helium gas through the vat of cleaning fluid entrains and removes the unstable ^{37}Ar atoms, which are swept into an external detector that looks for the x-rays.

To say that the extraction procedure needs to be efficient is a severe understatement. Only a single atom of ^{37}Ar is produced every two days in a 615-ton detector housing some 2×10^{30} chlorine nuclei! Yet exhaustive tests of the ^{37}Ar extraction process have proved it is not only efficient, but also reliable. After twenty-seven years of operation in the Homestake Mine, the chlorine experiment has measured an average neutrino flux that falls a factor of 3 below the predictions of the standard solar model.

The Kamiokande Experiment. The discrepancy between the measured and predicted solar-neutrino flux gave rise to much speculation about its origin, including exotic solutions such as neutrino oscillations. It also motivated the development of new solar-neutrino experiments.

First to come online of the "second-generation" experiments was Kamiokande. That experiment, situated 1 kilometer underground in the Kamioka Mine in the Japanese Alps, used a detector that was originally constructed to search for proton decay. Although the proton is a stable particle in the Standard Model of elementary particles, Grand Unified Theories predict that the proton would decay, albeit with an extremely long half-life.

But Kamiokande revealed no proton decays, and given the persistence of the solar-neutrino deficit measured in the chlorine experiment, the detector was upgraded to increase its sensitivity to the solar-neutrino signal.

The Kamiokande detector—a huge, cylindrical tank filled with 3,000 tons of purified water (H_2O)—is a Cerenkov detector. Neutrinos are detected in real time after they undergo elastic scattering with electrons in the water target:

$$\nu_e + e^- \rightarrow \nu_e + e^- . \qquad (2)$$
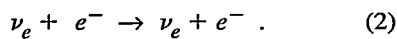
The neutrino imparts energy to the electron, which streaks through the water at relativistic speeds. The Cerenkov radiation emitted by that speedy electron was detected by an array of photomultiplier tubes surrounding the water target. Reconstructing the event allowed determining the energy and direction of the neutrino. Of all the first- and second-generation neutrino experiments, only Kamiokande could determine that the detected neutrinos indeed originated in the Sun.

The Kamiokande experiment could detect neutrinos with energies greater than 7 MeV. (The signals from lower-energy neutrinos were overwhelmed by detector background signals.) Kamiokande therefore provided a measurement of only the high-energy portion of the ^8B solar-neutrino flux (see Figure 2). After about 2,000 days of data acquisition, the Kamiokande collaboration reported results that were a factor of 2 below the predictions of the standard solar model. Hence, like the chlorine experiment, Kamiokande witnessed a significant deficit of solar neutrinos by comparison with solar-model predictions.

How does one reconcile this nagging discrepancy between theory and experiment? As mentioned earlier, the intensity of the ^7Be and ^8B neutrino fluxes depends delicately on the details of astrophysical models. Indeed, the core temperature of the Sun need only

be a wee bit lower in order to reduce the ^8B flux prediction by a factor of 2. Furthermore, the ^8B branch of the solar-neutrino spectrum constitutes a meager 0.01 percent of the total integrated neutrino flux from the Sun. Hence, one is inclined to simply ignore the small departure from theory and to appreciate a remarkable achievement: The solar model has predicted correctly (albeit within a factor of about 2) the intensity of a 0.01 percent branch of a solar-neutrino signal that varies as the 25th power of the core temperature! This fact alone is a clear indication that the basic ingredients of the astrophysical models are sound and that the models of the Sun and other main-sequence stars are essentially correct.

Nonetheless, physicists are a tenacious breed. Confident of their models and their experimental results, they speculated that the explanation behind the "missing" neutrinos lay fundamentally in the properties of the neutrino, rather than in some misunderstanding about astrophysics. And while the strong dependence of the ^8B flux on temperature left significant room to question the solar models, the fundamental nature of the neutrino deficit became even more compelling with data from two experiments that measured the dominant pp neutrino flux.

Gallium Experiments. SAGE (for Soviet-American, later Russian-American, gallium experiment) and GALLEX (gallium experiment) were new radiochemical experiments similar in nature to the chlorine experiment. However, they employed gallium as the neutrino target, which extended the energy sensitivity of the detector down to the energy of the pp neutrinos. The experiments were based on the reaction

$$\nu_e + {}^{71}Ga \rightarrow {}^{71}Ge + e^- . \qquad (3)$$
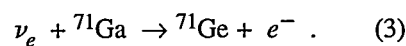
This reaction has an energy threshold of only 0.2332 MeV. As shown in Figure 2, SAGE and GALLEX were sensitive to all components of the solar-neutrino spectrum. SAGE, for which

Los Alamos National Laboratory was the lead American institution, is described in detail in the box "The Russian-American Gallium Experiment" on page 152.

The detection and measurement of the pp flux were something of a triumph for physicists. They represented the first experimental verification that the sun is indeed powered by thermonuclear fusion and that the power generated by the sun derives mostly from the pp fusion reaction. Again, theorists were led to conclude that the basic ingredients describing how the sun shines were understood and correctly implemented in astrophysical models.

However, the data from both SAGE and GALLEX confirmed the solar-neutrino deficit. Approximately half of the expected flux was observed. Those results significantly reshaped our understanding of the solar-neutrino problem.

The Modern Solar-Neutrino Puzzle

Data from the four pioneering experiments—chlorine, Kamiokande, SAGE, and GALLEX—provided information essentially across the entire solar-neutrino spectrum, from the low-energy and dominant pp flux up to and including the high-energy, but much less intense, ^8B flux. The experimental results and theoretical predictions are summarized in Table I. In all cases, the experimental results fall significantly below the predictions of the standard solar model. However, as indicated by the analysis outlined in Table II, the problem is even more perplexing.

Because each experiment is sensitive to different but overlapping regions of the solar-neutrino spectrum, the individual contributions to the total neutrino flux of the pp, ^7Be, and ^8B neutrinos can be estimated. All that is required is an analytical procedure that simultaneously takes into account all the experimental results and the solar-luminosity constraint. The intriguing result is

Table I. Summary of Pioneering Solar-Neutrino Experiments

	SAGE + GALLEX	Chlorine	Kamiokande
Target Material	^{71}Ga	^{37}Cl	H_2O
Reaction	$\nu_e + {^{71}}\text{Ga} \rightarrow {^{71}}\text{Ge} + e^-$	$\nu_e + {^{37}}\text{Cl} \rightarrow {^{37}}\text{Ar} + e^-$	$\nu + e^- \rightarrow \nu + e^-$
Detection Method	Radiochemical	Radiochemical	Cerenkov
Detection Threshold	0.234 MeV	0.814 MeV	7.0 MeV
Neutrinos Detected	All	^{7}Be and ^{8}B	^{8}B
Predicted Rate	132 ± 7 SNU*	9 ± 1 SNU	5.7 ± 0.8 flux units**
Observed Rate	74 ± 8 SNU	2.5 ± 0.2 SNU	2.9 ± 0.4 flux units

*1 SNU = 10^{-36} captures per target atom per second.
**In units of 10^6 neutrinos per square centimeter per second.

Each column summarizes an experiment and compares the predicted rate of neutrino interactions (based on the Bahcall-Pinsonneault standard solar model) to the observed rate. The radiochemical experiments report their results in SNU, a convenient unit that facilitates comparison between experiments. Kamiokande reports results in flux units. Every experiment shows a significant deficit in the observed versus the predicted rate.

Table II. Breakdown of the Predicted Rate by Neutrino-Producing Reaction

Neutrino Reaction	SAGE + GALLEX	Chlorine	Kamiokande
pp	70 SNU	0 SNU	0
pep	3 SNU	0.2 SNU	0
^{7}Be	38 SNU	1.2 SNU	0
^{8}B	16 SNU	7.4 SNU	5.7
CNO	10 SNU	0.5 SNU	0
Total Predicted Rate	132 ± 7 SNU	9 ± 1 SNU	5.7 ± 0.8
Observed Rate	74 ± 8 SNU	2.5 ± 0.2 SNU	2.9 ± 0.4

Based on the standard solar model, the total predicted rate of neutrino events can be broken down into contributions from each of the neutrino-producing reactions in the Sun. This information is listed in each column (rounded to the nearest SNU) and is displayed as a bar graph. (The bars corresponding to the total predicted rate have been normalized to 1.) Each colored segment within a bar corresponds to a specific reaction. Kamiokande observed approximately half of the expected flux of ^{8}B neutrinos. All the neutrinos detected by the chlorine experiment can likewise come from the ^{8}B reaction. The solar luminosity essentially fixes the rate of *pp* neutrinos that SAGE and GALLEX must see. Those experiments are consistent with an observation of the full *pp* flux plus some of the ^{8}B flux. Taken together, the experiments indicate that the solar-neutrino deficit results from a lack of intermediate-energy (CNO, ^{7}Be, and *pep*) neutrinos.
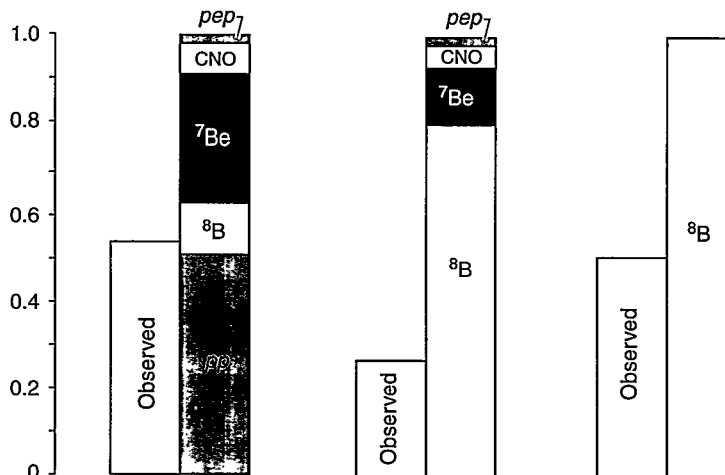
Figure 3. The Modern Solar-Neutrino Problem

One can deduce how the theoretical neutrino flux needs to be distorted in order to match the experimental results. In their analysis, Hata and Langacker (1994) constructed an arbitrary solar model in which the neutrino fluxes are allowed to vary freely instead of being tied to nuclear physics or to astrophysics. The only constraint is the one imposed by the solar luminosity, namely, that the sum of the pp, ^7Be, and CNO fluxes roughly equals 6.57×10^{10} neutrinos per square centimeter per second (the total neutrino flux). The model is then "fit" to the combined data from all experiments.

Source of Neutrinos	Φ/Φ^{SSM}
pp	1
^7Be	0
^8B	0.4

The model that best fits the data is one in which the pp flux is identical with the standard-solar-model (SSM) prediction, the ^7Be flux is nearly absent, and the ^8B flux is only 40 percent of the SSM prediction. These results are presented in the table (left) as the ratio of Φ, the flux derived from the combined fit, to Φ^{SSM}, which is the neutrino flux predicted by the SSM.
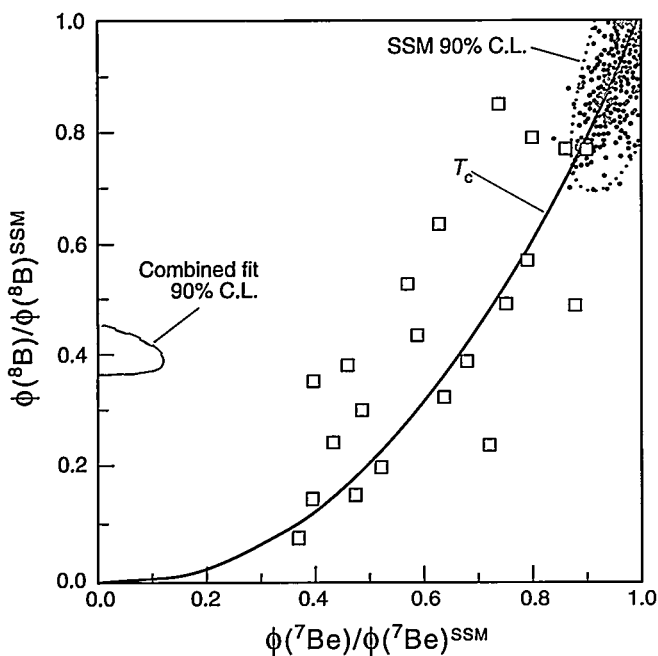
The 90 percent confidence level for the combined fit is shown in blue on this graph of ^8B flux versus ^7Be flux (each normalized to the SSM predictions). The 90 percent confidence level for the Bahcall-Pinsonneault SSM is shown at the upper right-hand corner. Filling that contour are the results of 1,000 Monte Carlo simulations (green dots) that vary the parameters of the SSM. The square markers indicate the results of numerous nonstandard solar models, which include, for example, variations in reaction cross sections, reduced heavy-element abundances, reduced opacity models, and even weakly interacting massive particles. Most of the models call for a power law relation between the ^8B and ^7Be fluxes (the curve labeled T_c). As the figure shows, the SSM and all nonstandard models are completely at odds with the best fit to the combined experimental results.

shown in Figure 3. Compared with the solar-model predictions, the pp neutrino flux, with a maximum neutrino energy of 0.42 MeV, seems to be present in full strength. The intermediate-energy ^7Be neutrinos, however, seem to be missing entirely, while only 40 percent of the high-energy ^8B neutrinos are observed.

This energy-dependent suppression of the solar-neutrino spectrum establishes what we now refer to as the *modern solar-neutrino problem*. It is particularly puzzling given the apparent lack of ^7Be neutrinos. At a glance, this might imply that ^7Be is not being produced in the sun. But those nuclei are needed to produce ^8B (refer to

Figure 1). Hence, if there are no ^7Be neutrinos, why are *any* ^8B neutrinos observed? While modifications to the solar models have been attempted by many authors, it appears extremely difficult to render an astrophysical explanation that would solve this puzzle. As seen in Figure 3, no model has successfully reduced the ^7Be flux without reducing the ^8B flux even more!

However, this pattern for the solar-neutrino spectrum is perfectly explained by the mechanism of matter-enhanced neutrino oscillations, or the MSW effect. (See the article "MSW" on page 156.) MSW suggests that the probability for neutrino oscillations to occur *in vacuo* can be augmented in an energy-

dependent, resonant fashion when neutrinos travel through dense matter. The muon or tau neutrinos would not be detected in the existing experiments on Earth, and hence a deficit would be seen in the solar-neutrino flux. For suitable choices of neutrino masses and mixing angles, experiments would measure the full, predicted flux of pp neutrinos, the ^7Be flux would be highly suppressed, and the measured flux of ^8B neutrinos would be reduced to 40 percent! (See Figure 4.)

Have three decades of solar-neutrino research culminated in the discovery of neutrino mass? Our interpretation of the modern solar-neutrino problem relies upon our confidence that the standard
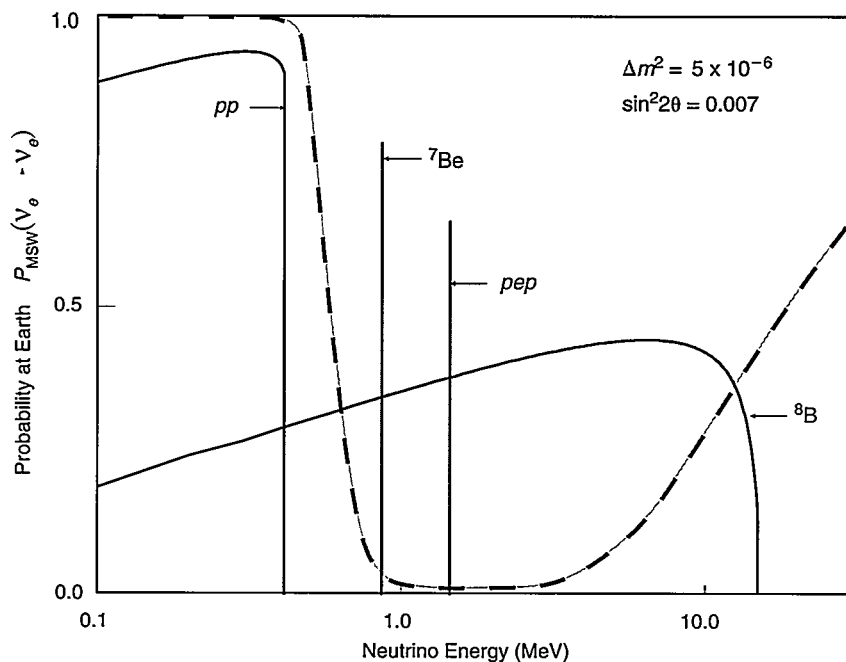
Figure 4. Energy-Dependent Survival Probability for Solar Neutrinos
The MSW effect predicts that the energy spectrum of the Sun's electron neutrinos
would be distorted from standard-solar-model predictions. Because of their charged-
current interactions with electrons, electron neutrinos can acquire an effective mass
when they pass through dense matter. If that effective mass becomes as large as the
intrinsic mass of a muon or tau neutrino, the electron neutrino can resonantly trans-
form into another flavor. The spectrum becomes distorted. The MSW survival
probability $P_{MSW}(\nu_e \to \nu_e)$ that an electron neutrino remains an electron neutrino
depends on the neutrino mass difference (Δm^2), the mixing angle θ between the
different neutrino flavors, the neutrino energy, the electron density in the material,
and the strength of the interaction between electron neutrinos and electrons. Given
the Sun's density profile, the chosen values for Δm^2 and $\sin^2 2\theta$ will yield a survival
probability that distorts the predicted solar-neutrino energy spectrum for best
agreement with the measured flux. The result is the red curve, which has been
superimposed over a simplified picture of the solar-neutrino spectrum. The pp flux
is unaffected, but the ^7Be and pep fluxes have almost no probability of surviving.
Roughly 40 percent of the ^8B neutrino flux survives.

solar model correctly predicts the solar-
neutrino spectrum and that the existing
experimental results are accurate. Given
the implications of massive neutrinos
for elementary particle physics,
astrophysics, and cosmology, it is
of paramount importance to confirm
the solar-neutrino problem and to test
the hypothesis of neutrino oscillations.
Thus, while theorists continue to
analyze and refine the solar models,
experimentalists are moving forward
with the next generation of experiments.

The super-Kamiokande experiment,

a 50,000-ton water-filled Cerenkov
detector, is currently up and running.
Because of its immense size, this detec-
tor accumulates in one day more data
than its predecessor Kamiokande was
capable of providing in one month!
It has collected a significant amount
of data on the high-energy region of
the ^8B neutrino flux and has confirmed
the results of Kamiokande.

Complementing that experiment is
BOREXINO, an experiment that is still
in its planning stages. The proposal
calls for a detector to be situated in the

Gran Sasso tunnel in Italy. Like super-
Kamiokande, BOREXINO exploits the
scattering of neutrinos from electrons in
the target material, but unlike super-
Kamiokande, it will use a liquid scin-
tillator instead of water as the neutrino
target. This allows for a much lower
energy threshold so that the experiment
will be highly sensitive to the ^7Be neu-
trinos. Because of the inherent energy
resolution of its detector, BOREXINO
should be able to focus on and isolate
the ^7Be neutrino flux and thus allow
scientists to deduce, independently,
whether that branch is indeed missing
from the solar-neutrino spectrum.

But it is fitting at this point to
reemphasize that the Sun is energetic
enough to produce only electron neutri-
nos and that the pioneering experiments
were sensitive *only* to electron neutrinos.
Deductions about neutrino oscillations
came about only because the measured
electron neutrino flux was found lacking
when compared with the predictions of
the standard solar model.

The ideal experiment would not
have to rely on a model to allow data
interpretation. Such an experiment
would independently measure the
electron neutrino flux and the *total*
neutrino flux (electron, muon, and tau
neutrinos). Independent measurement of
the latter is perhaps the "Holy Grail" of
solar-neutrino experiments. If electron
neutrinos are experiencing flavor
transitions into other states, then the
electron neutrino flux would be
observed to be lower than the total flux,
and the neutrino oscillation hypothesis
would be tested in a *model-independent*
fashion. Such an experiment forms
the motivation for establishing the
Sudbury Neutrino Observatory.[1]

[1]In principle, super-Kamiokande is also sensitive
to all neutrino flavors via the neutral-current
channel of elastic scattering. The neutral- and
charged-current signals, however, are identical
and cannot be distinguished from each other.
Experimenters cannot analyze the data and
deduce that they have seen muon or tau neutrinos
from the Sun, unless they compare their mea-
sured flux with one predicted by the solar model.

The Sudbury Neutrino Observatory (SNO)

SNO is a next-generation, real-time experiment that is designed to make independent measurements of (1) the flux and energy spectrum of ^8B electron neutrinos reaching the earth and (2) the *total* integrated flux of all ^8B neutrinos that reach the earth. These goals can be met because the water Cerenkov detector that is at the heart of SNO will be filled with a unique neutrino target consisting entirely of "heavy" water.

A heavy-water molecule has two deuterium atoms bonded to an oxygen atom (D_2O rather than H_2O). The deuterium nucleus—the deuteron—is a heavy isotope of hydrogen that consists of a bound proton and neutron. Thus, heavy water is chemically identical to ordinary "light" water (H_2O), and the SNO detector functions very much like the light-water Cerenkov detector used by Kamiokande and, later, super-Kamiokande. But the signal from a light-water Cerenkov detector derives solely from the elastic scattering of neutrinos with electrons. The heavy water in the SNO detector also allows for neutrino interactions with the nucleons making up the deuterium nucleus. And crucial to the SNO experiment, deuterium responds in different ways to charged- and neutral-current interactions (see Figure 5).

When the deuteron (D) interacts with electron neutrinos through the charged-current exchange of a W^+ boson, the neutron transforms into a proton, and a relativistic electron is emitted. The newly created nucleus contains two protons. These repel each other and break the nucleus apart:

$$\nu_e + D \rightarrow p + p + e^- . \qquad (4)$$
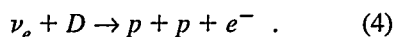
The protons do not recoil with sufficient energy to create a signal in the detector, but the electron produces Cerenkov radiation as it zips through the heavy water. The energy and direction of the electron neutrino can be

(a) Charged-Current Interaction

$$\nu_e + D \rightarrow p + p + e^-$$
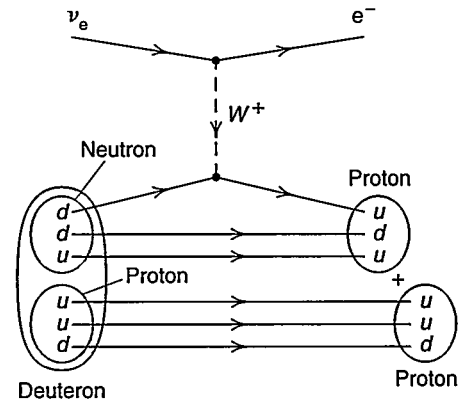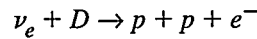
(b) Neutral-Current Interaction

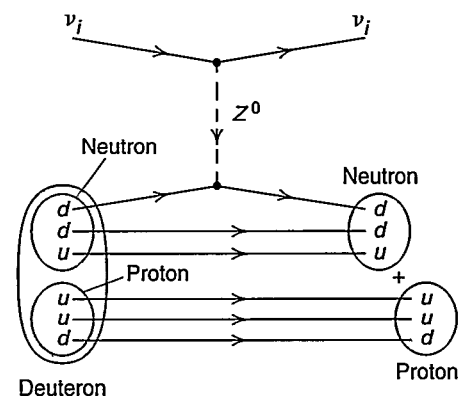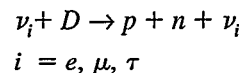$$\nu_i + D \rightarrow p + n + \nu_i$$
$$i = e, \mu, \tau$$

Figure 5. Neutrino Interactions in SNO
(a) The charged-current weak interaction in SNO proceeds only with electron neutrinos ν_e. The neutrino is transformed into an electron e^- as it exchanges a W^+ boson with one of the two down quarks d making up the neutron (udd quark combination). The quark is transformed into an up quark u, thus creating a proton (uud). The proton originally present in the deuteron does not participate in the interaction; it is merely a spectator. The unstable diproton system instantly breaks apart, so that the deuteron appears to disintegrate into two free protons and a relativistic electron. That electron is the signature of the reaction. (b) The neutral-current interaction can proceed with a neutrino of any flavor. The neutrino scatters from one of the quarks in either nucleon (proton or neutron) through the exchange of a Z^0 boson. Energy is transferred to the entire nucleon. If the energy transfer is greater than the 2.22 MeV nuclear binding energy, the deuteron breaks apart. The unbound neutron is the signature of the neutral-current interaction.

extracted from that Cerenkov signal.

The neutral-current interaction, however, causes the deuteron to disintegrate without any change in particle identity. An exchange of a Z^0 boson transfers energy into the deuteron, which breaks into a proton and a neutron:

$$\nu + D \rightarrow \nu + p + n . \qquad (5)$$
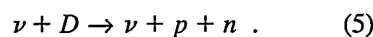
Reaction (5) occurs with equal probability for any flavor of neutrino whose

energy is above the 2.22-MeV binding energy of the deuteron. Detecting the free neutron is the key to observing this reaction, and techniques for extracting that signal will be discussed later. The main point to appreciate is that SNO can detect all neutrinos (regardless of flavor) and the electron neutrinos in two independent measurements. A comparison of the two fluxes will provide a definitive test of the

Figure 6. The Sudbury Neutrino Observatory

The SNO detector sits in a large cavity that is over 2 kilometers underground in the Creighton Mine in Sudbury, Canada. The barrel-shaped cavity is 22 meters in diameter and 30 meters high. In its center lies an acrylic bottle that will contain about 1,000 tons of 99.92 percent isotopically enriched D_2O. The 12-meter-diameter acrylic sphere at the bottom of the bottle, the largest sphere of its kind in the world, is constructed of 122 panels, each 5 centimeters thick, that are machined and thermoformed before being assembled underground. Surrounding the sphere will be a geodesic frame anchoring an array of 9,800 photomultiplier tubes (PMTs). The 20-centimeter-diameter PMTs will be equidistantly spaced, 2.5 meters apart, outside the acrylic vessel. Filling the remainder of the cavity will be about 7,500 tons of ultrapure H_2O, which shields the D_2O volume from radioactive backgrounds while simultaneously providing mechanical support for the acrylic vessel. Electron neutrinos interact in the detector to produce relativistic electrons. The electrons create Cerenkov radiation that is detected by the PMTs. Based on the Cerenkov photons' time of arrival at the PMTs and the number of PMTs that are triggered, the location, energy, and direction of the electron can be determined. All neutrino flavors can create free neutrons in the D_2O by neutral-current interactions. The neutrons can be detected through the Cerenkov light produced by neutron capture on chlorine dissolved in the D_2O or, otherwise, by a secondary detector that would sit inside the acrylic bottle.
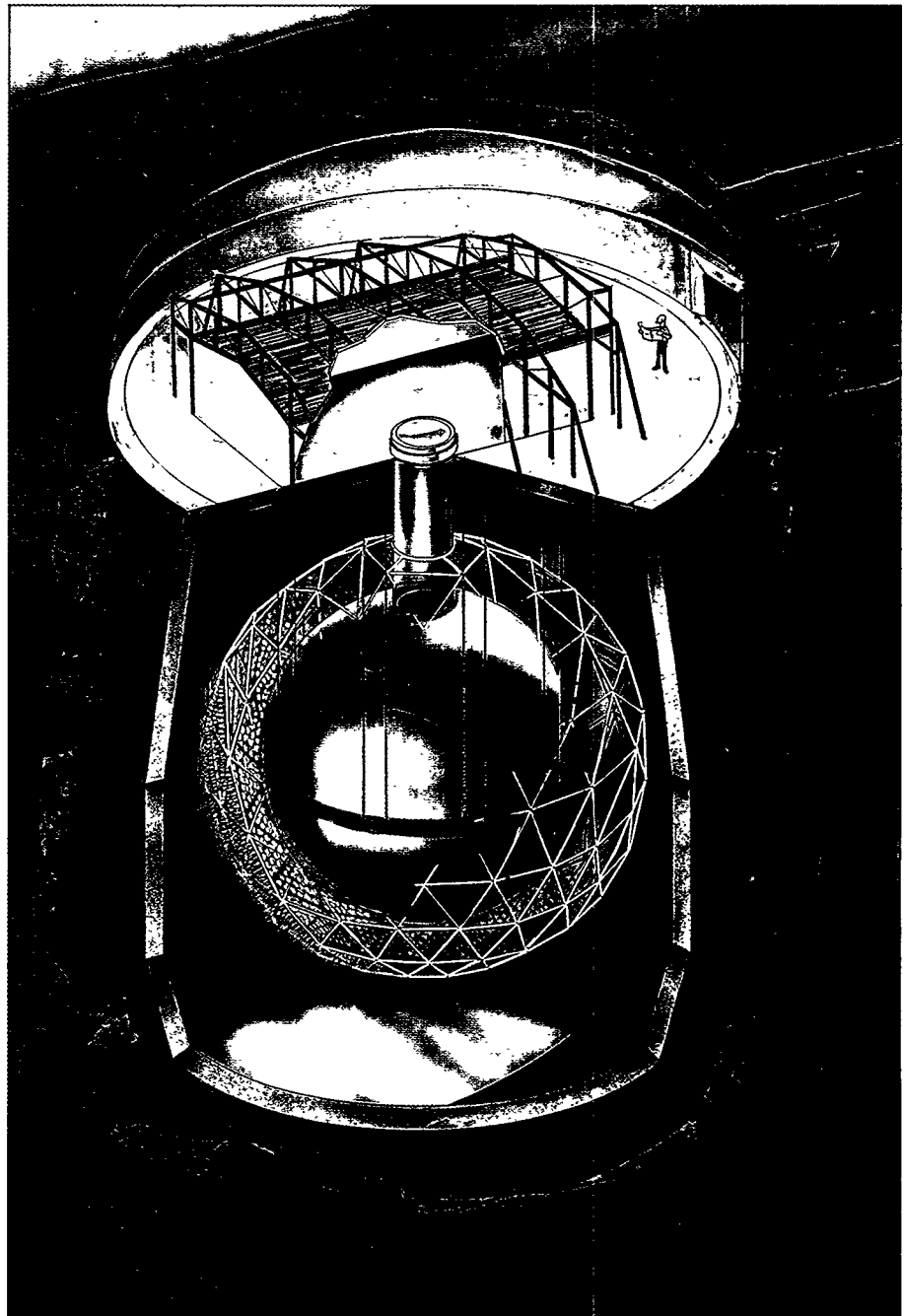
neutrino oscillation hypothesis.

The idea of using heavy water as a neutrino target was first proposed in 1984 by H. H. Chen. Upon visiting the Creighton Mine in Sudbury, Canada, to discuss a proton decay experiment, Chen became aware that the Atomic Energy Commission of Canada had large stockpiles of heavy water. He subsequently pursued the possibility of borrowing about 1,000 tons. The SNO collaboration was formed shortly there-

after, and a feasibility study started in 1985 to evaluate the capabilities and practicality of building a heavy-water Cerenkov detector.

In 1986, an exploratory site was located wherein a cavity 20 meters in diameter could be constructed 6,800 feet underground in the Creighton Mine. A detailed proposal for SNO was reviewed in June 1988 by an international scientific and technical review committee that recommended that SNO be

approved and funded as proposed. The SNO project is currently under construction by collaborators from twelve institutions in Canada, the United States, and the United Kingdom.

As shown in Figure 6, the 1,000 tons of 99.92 percent isotopically enriched heavy water will fill a huge acrylic bottle that is the centerpiece of the SNO detector. Surrounding the bottle will be 9,800 photomultiplier tubes that will detect the feeble Cerenkov light.

Approximately 7,500 tons of purified light water will encase the bottle and phototubes. That light-water jacket is needed to shield the detector from radioactive emissions emanating from the rock surrounding SNO.

The Charged-Current Spectrum in SNO. Measuring the Cerenkov spectrum due to the charged-current interaction of ^{8}B electron neutrinos shown in Reaction (4) is one of the primary goals of SNO. Should the neutral- to charged-current ratio indicate neutrino oscillations, the shape of the charged-current spectrum could be used to probe different solutions to the solar-neutrino problem. For example, the MSW effect predicts a depletion in the flux of lower-energy ^{8}B neutrinos, and this reduced flux would be mostly evident as a change in the shape of the spectrum between about 5 and 8 MeV, as shown in Figure 7. SNO's detection ability and sensitivity to the charged-current signal have been assessed with computer simulations that predict the response of the detector to that signal and various anticipated background signals.

An example of such a simulation is shown in Figure 8. Below about 4 MeV, the detector is recording Cerenkov light that is due mostly to background processes (the "Cerenkov background wall"). Uranium and thorium atoms, which will unavoidably contaminate the heavy water and the detector materials, decay and produce energetic beta particles and gamma rays. These emissions create Cerenkov light when they streak through the heavy water. Signals due to neutrino events cannot be discerned beneath this wall of background light, and thus SNO is only expected to be sensitive to neutrinos with energies greater than about 5 MeV.

It is also evident from Figure 8 that between about 5 and 8 MeV, the summed Cerenkov spectrum derives from a complex overlap of different signals. The charged-current spectrum, which extends all the way to about 14 MeV, peaks in that region. But the
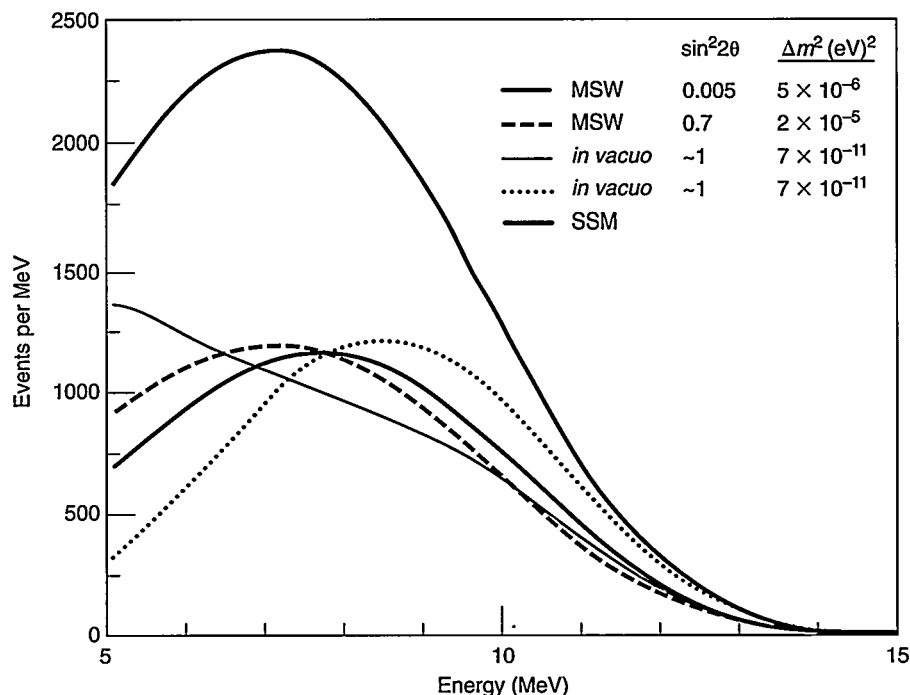

Figure 7. Theoretical Distortions in the Charged-Current Spectrum
By assuming a ^{8}B neutrino spectrum, one can simulate what SNO would record for the charged-current (electron neutrino) Cerenkov spectrum. The black curve results from the standard-solar-model (SSM) spectrum, whereas the other curves are the result of distorting the ^{8}B spectrum either through *in vacuo* neutrino oscillations (red curves) or the MSW effect (blue curves). The amount of distortion depends on the amplitude and wavelength of the oscillations, which are characterized by the amount of mixing between neutrino flavors (the parameter $\sin^2 2\theta$) and the mass difference between neutrino mass states (the parameter Δm^2), respectively. The red curves reflect two of the five "just so" solutions for *in vacuo* neutrino oscillations, labeled as such because the large mixing angles and tiny mass differences are just right to make the oscillation length match the earth's orbit. The two blue curves derive from two different MSW solutions that are consistent with the existing data. The solid curve results when one assumes that electron neutrinos become muon neutrinos over a density range that is short compared with the neutrino oscillation length (nonadiabatic MSW solution), whereas the dashed curve results from a theory that assumes essentially the opposite (adiabatic, or large-angle, solution). The most favored solution to the solar-neutrino problem is the nonadiabatic MSW solution.

neutral-current spectrum and neutrino elastic-scattering spectrum are also present. Whereas detecting these latter signals is one of the design goals for SNO, in the context of isolating and measuring the charged-current spectrum, the signals represent complicating backgrounds.

The simulation shows the importance of maintaining an ultraclean detector environment in order to minimize the Cerenkov wall, especially in the critical region between 5 and

8 MeV. (See the box "Nothing to Dust: The Meaning of Clean" on page 149.) Although ensuring the radiopurity of construction materials has been a major focus in this project, the background levels in the light-water jacket and in the heavy water will be monitored by a variety of techniques. In addition, several calibration sources will establish the optical properties of the SNO detector and its response to electrons, gamma rays, and neutrons. These sources will be inserted in the·

Figure 8. Monte Carlo Simulation of the Charged-Current Spectrum and Backgrounds in SNO
Many processes produce Cerenkov light in the SNO detector, and the detector output (black line) is really the sum of many spectra. The charged-current spectrum is shown in blue, and all other spectra can be considered as backgrounds. The red spectrum results when radioactive impurities in the D_2O and the detector components decay. The beta and gamma emissions create the Cerenkov light. This spectrum completely dominates the detector response below about 4 MeV (the "Cerenkov background wall"). The dark green spectrum results when a free neutron, produced through either the neutral-current interaction or the gamma-ray-induced photodisintegration of deuterium, is captured by a deuterium nucleus. The neutron capture results in a Cerenkov-producing gamma ray. The light green spectrum results when neutrinos undergo elastic scattering with electrons, a process shown in Reaction (2) in the main text. Energy is transferred to the electron, which zips through the heavy water. The components of the summed spectra are particularly difficult to sort out between about 5 and 8 MeV, precisely the region that is of most interest.
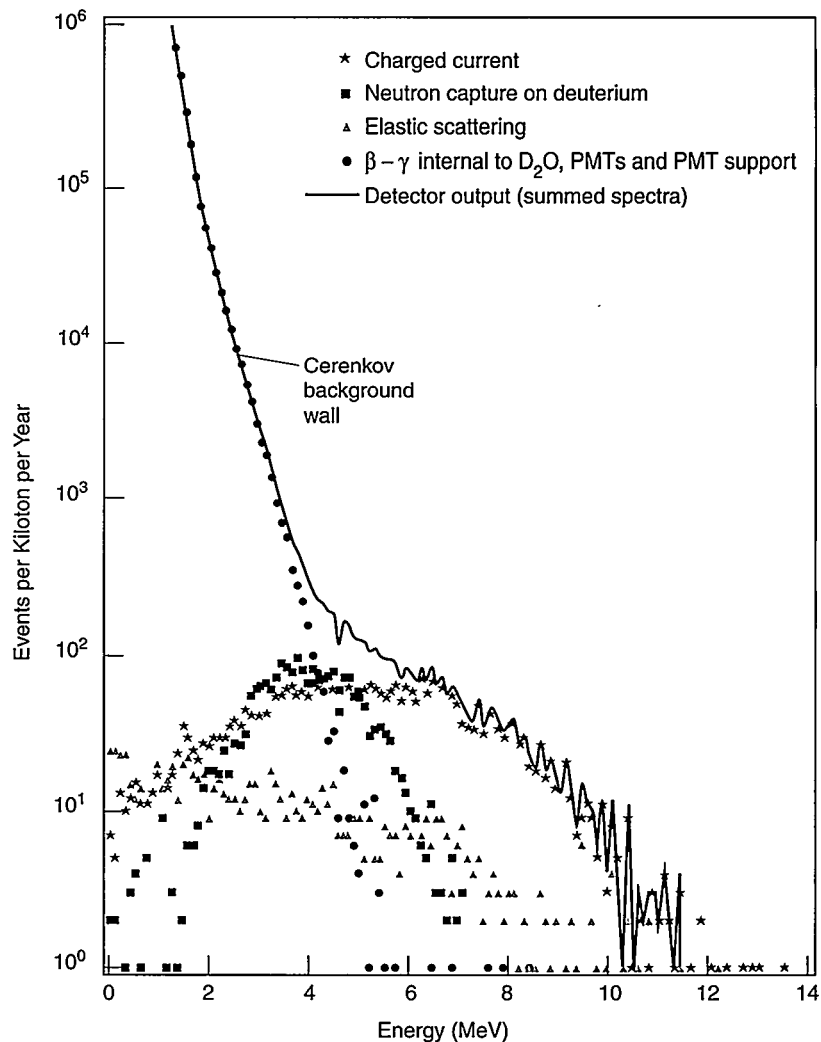
heavy-water vessel and will mimic a wide range of natural sources.

Neutral-Current Detection. The neutron liberated by the disintegration of the deuteron—Reaction (5)—leads to the all-important neutral-current signal (from which one infers the total solar-neutrino flux). The free neutron can be captured by a deuterium nucleus, which then emits a 6.25-MeV gamma ray that creates a shower of Cerenkov radiation in the detector. The spectrum is shown in Figure 8 (dark green squares).

However, the signal produced by neutron capture on deuterium would be very difficult to observe. As shown in the figure, the spectrum lies substantially beneath the Cerenkov background wall, which means that the neutral-current

signal would be overwhelmed by background signals. One way around this problem is to "boost" the neutral-current signal toward higher energies. Neutron capture on ^{35}Cl yields an 8.6-MeV gamma ray, which produces Cerenkov light that can be safely discriminated against background. Dissolving about 2.5 tons of magnesium salt ($MgCl_2$) into the heavy water should allow this detection method.

But there is still a complication. Because both the neutral- and charged-current interactions produce Cerenkov light, the two signals become part of the total Cerenkov signal that is recorded by SNO. In effect, the two signals become backgrounds to each other. To disentangle the signals, the detector would have to operate first

with and then without salt. A subtraction of one data set from the other would allow separation of the charged-current and neutral-current signals. As an alternative, the data could be analyzed with sophisticated pattern-recognition techniques. These could help discriminate a neutron event from other Cerenkov signals and would obviate the need for data subtractions.

It is clearly desirable to have a direct way of distinguishing between neutral- and charged-current events in real time. But a direct detection method can only be achieved if the neutrons produced by the neutral-current interaction are observed by means other than Cerenkov light. To this end, a research and development program started at Los Alamos several years ago. Its goal

148

Nothing to Dust: The Meaning of Clean

The standard solar model predicts that electron neutrinos will produce about 30 charged-current events per day in the SNO detector. According to the interpretation of the existing solar-neutrino results, this number will be reduced by about a factor of 2 to 3. Similarly, one expects about 14 neutrons to be produced per day through the neutral-current interaction. Relatively speaking, those estimated rates make SNO a "high-rate" solar-neutrino detector, but obviously, neutrino events are sufficiently rare that great strides must be taken to ensure an extraordinarily low background environment.

Any Cerenkov radiation that does not originate from a neutrino event is a background signal. The background sources could be cosmic rays or the energetic beta and gamma rays coming from the decay of radioactive elements. The cosmic-ray background is eliminated simply because SNO is buried underground with an overburden of 6,800 feet of rock. Also, emissions from radioactive elements in the norite rock surrounding the detector are eliminated by the 7,500 tons of light water engulfing the acrylic heavy-water (D_2O) bottle. All this leaves the dominant source of background to be one from within, that is, from the radiation emitted by radionuclide impurities present in the materials used to construct the detector.

All materials naturally contain small quantities of radionuclides. Unfortunately, those quantities are in general many orders of magnitude larger than what can be tolerated in a solar-neutrino experiment. For example, if the long-lived isotopes thorium-232 (^{232}Th) and uranium-238 (^{238}U) were present in even minute quantities, their beta and gamma activity would produce enough Cerenkov radiation to completely mask the electron neutrino signal. Also, the decay of thallium-208 and bismuth-214, which lie at the bottom of the ^{232}Th and ^{238}U decay chains, respectively, yields gamma rays above the binding energy of the deuteron. The gamma rays can cause the deuteron to photodisintegrate and liberate a neutron that is indistinguishable from neutrons produced via the all-important neutral-current interaction. Consequently, any material considered for use in constructing the SNO detector must be chosen with extreme care. An extensive research program was initiated to identify appropriate construction materials and to measure the intrinsic radioactive-impurity levels.

The most stringent requirement for purity falls upon the heavy water because there is so much of it. Specifically, thorium and uranium levels must be reduced to parts in 10^{14} by weight for the ensuing backgrounds not to exceed about 10 percent of the expected solar-neutrino signal. In other words, the 1,000 tons of heavy water cannot contain more than about 10 *micrograms* of heavy radioactive isotopes! Impurity levels in the acrylic bottle are less restrictive because of the bottle's smaller mass and must be reduced to parts in 10^{12} by weight. Fortunately, because acrylic turns out to be an intrinsically pure material with respect to the nasty thorium and uranium, it can meet the strict requirements of the SNO detector. The photomultipliers for detecting Cerenkov light, the photomultiplier support structure, and the cables—all poised roughly 2.5 meters from the acrylic vessel—must also meet strict requirements for radiopurity. The light-water shield effectively attenuates most of the radioemissions before they enter the D_2O target. Nonetheless, the photomultipliers are constructed from a low-radioactivity glass containing thorium and uranium at levels 10 times lower than those of standard glass.

All these purity constraints would be moot if, during the construction phase, dust and impurities from the outside world were reintroduced into the materials. For this reason, the site where SNO is under construction—an enormous cavity more than 1 mile underground—has been turned into a giant clean room with surface-deposition rates of dust particles kept to about 1 microgram per square centimeter per month. Paradoxically, a normally filthy environment now houses one of the world's cleanest rooms.

was to design a discrete, ultralow-background neutron detector that could be deployed inside the heavy-water vessel. This secondary detector would have the capability of independently detecting neutrons while working simultaneously with the main SNO detector. After several years of development, such a detector is entering the stage of full-scale construction by SNO collaborators at Los Alamos, the University of Washington, and

Lawrence Berkeley National Laboratory.

The Discrete Neutral-Current Detector. The discrete detector is an array of ^{3}He proportional counters. These are standard devices for detecting neutrons with high efficiency, and they are frequently used in nuclear and particle physics experiments. Each counter is made up of a cylindrical tube filled with a gas mixture containing ^{3}He. Neutrons easily penetrate the thin-

walled tube and are captured on ^{3}He:

$$^3He + n \rightarrow {}^3H + p . \qquad (6)$$
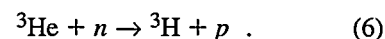
The reaction produces an energetic proton and triton, both of which lose energy by ionizing the gas molecules. The resulting cloud of charged ions is attracted to a thin wire strung down the center of the tube that is kept at an electrical potential of about 1,800 volts. Monitoring the high-voltage line for

Figure 9. The Neutral-Current Detector for SNO
The long, vertical tubes strung throughout the vessel containing heavy water are
^{3}He proportional counters. A neutron passing through the tube wall will trigger a
current pulse that is picked up by the signal lines snaking through the neck of the
acrylic bottle. The array of counters enables independent measurement of the neutrino
neutral-current interaction, even as the main detector measures the electron neutrino
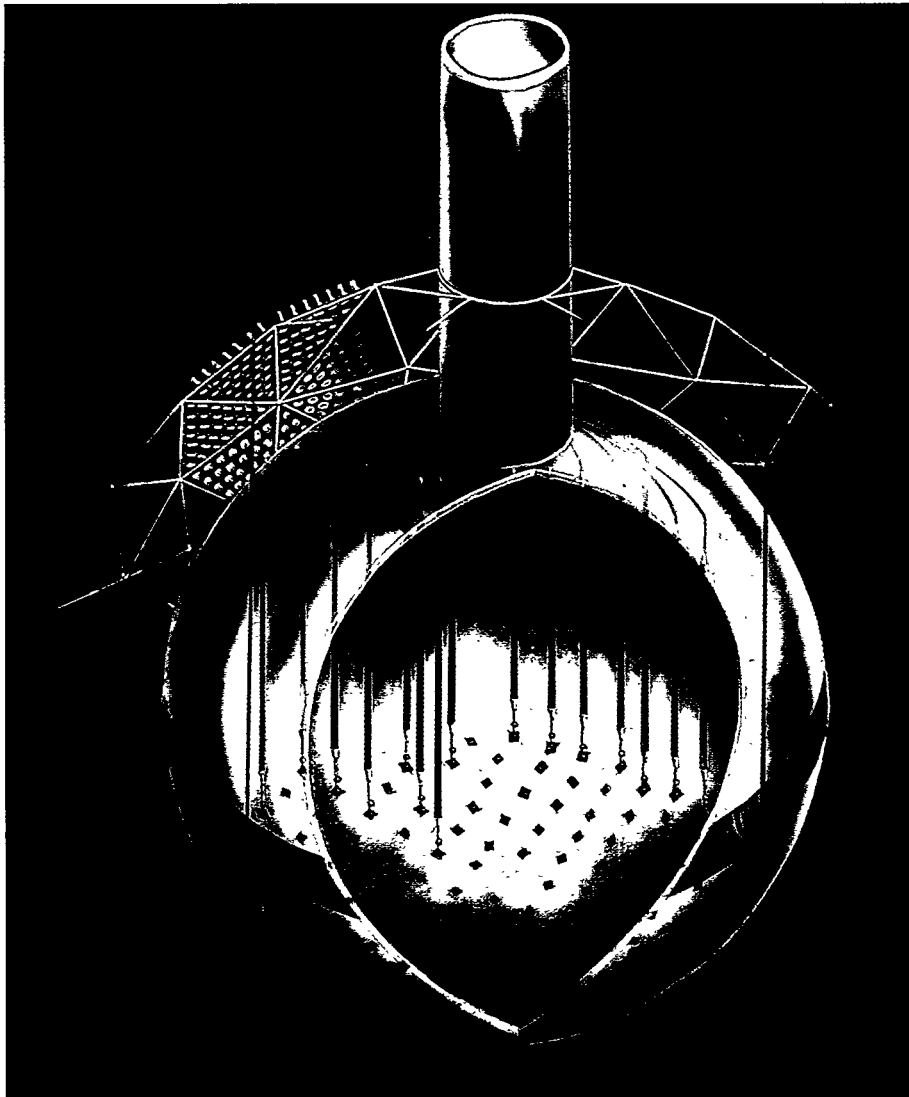charged-current interaction. The full array will use about 800 meters worth of propor-
tional counters and will include individual gas tubes up to 11 meters in length (very
large relative to conventional counters). While being immersed in water, the counters
must operate in a reliable and stable fashion for a period of up to 10 years.

than parts per trillion by weight.
Such levels are about 1,000 times
lower than those typically found in
commercially available construction
materials. Consequently, most of the
detector tubes have been fabricated
from ultrapure nickel by a special
chemical-vapor deposition process.
The inherent radioactive content of
these tubes is so low that the sensitivity
for extracting the neutron signal is
some 250 times greater than that of
any previously constructed, low-
background proportional counter.

Prototype detectors have been con-
structed and placed in an underground
test facility at the Waste Isolation Pilot
Plant near Carlsbad, New Mexico. The
underground tests have demonstrated
that the detectors can meet the cleanli-
ness requirements for the SNO
experiment. In addition, all the compo-
nents that enter into the construction
of the detector are assayed for their
radiopurity with sophisticated radio-
chemical techniques unique to Los
Alamos. An interesting application of
those techniques is described in the box
"A SNO Spinoff" on the facing page.

Summary and Outlook

Currently under construction in the
Creighton Mine in Sudbury, Canada,
by some 150 physicists, engineers,
and technicians, SNO has met many
triumphant milestones. The immense
cavity needed to house the detector
has been excavated, and the water-
purification systems are in place.
Construction of the heart of SNO is
well underway: the assembly of the
upper hemisphere of the heavy-water
acrylic vessel is finished, and the
upper portion of the photomultiplier
array is fully installed. Construction
of the lower half of the SNO
detector is proceeding. It is anticipated
that the detector will be ready for
operation and "first fill" toward
the end of 1997.

characteristic current pulses allows
detection and counting of the neutrons.
An artist's drawing of the array
anchored inside the acrylic bottle is
shown in Figure 9.

Designing ^{3}He proportional counters
for neutron detection is a long-standing
and well-practiced art, but several
challenging constraints had to be met

before the counters could realize
their potential as a secondary detector
in SNO. The most stringent constraint
was purity. Because the neutron signal
is predicted to be so low and because
the array is inside the main detector,
the bulk materials used to construct
the counters could not contain thorium
and uranium in concentrations greater

Full-scale construction of the
discrete neutral-current detector is now

A SNO Spinoff

Research into ultralow-background fabrication techniques has created a potentially interesting spinoff relevant to the microelectronics industry. Because computer chips are becoming extremely small, the decay of natural radioactive elements in the construction materials is sufficient to create "single-bit upsets" from one binary number to another. In an effort to solve this problem, the Weak Interactions Group at Los Alamos is currently working with industry to provide ultralow-background particle detectors for screening microelectronic components. In an interesting and unexpected development, something as esoteric as hunting neutrinos may lead to creating a useful tool in the "practical" world.

also underway and will continue in parallel with the construction of the main SNO detector over a period of about ten months. As independent proportional counters of the array are fabricated, they will be shipped to Sudbury and then stored underground for a period of six months. This "cooldown" period will allow any cosmogenically produced radioactivity to die out before the counters are deployed. After the SNO detector is filled with heavy water, a "shake-down" period will follow in order to assess the detector backgrounds, monitor the performance of the photomultiplier tubes, and fine-tune the many channels of electronics and data acquisition. When it is time to install the neutral-current detector, another delicate program will begin. Each counter must be installed through the top of the acrylic vessel with a remotely controlled submarine—so much effort to capture the elusive neutrino in its disappearing act!

But the neutrino has tempted and intrigued physicists for more than sixty years, ever since its existence was first postulated by Wolfgang Pauli. The present conundrum surrounding the missing solar neutrinos points to the possibility of very exciting physics. It may well be that, by the end of this century, the properties of the neutrino that at one time seemed undetectable may be revealed and may offer a long-awaited clue to some of the fundamental mysteries of the universe. ■

Further Reading

Bahcall, J. N. 1989. *Neutrino Astrophysics.* Cambridge, England: Cambridge University Press.

Bowles, T. J., et al. 1992. Neutral-Current Detection in the Sudbury Neutrino Observatory. Los Alamos National Laboratory proposal FIN-94-ER-E324.

Doe, P. J., et al. 1995. Construction of an Array of Neutral-Current Detectors for the Sudbury Neutrino Observatory. Sudbury Neutrino Observatory proposal SNO-STR-95-023.

Ewan, G. T., et al. 1992. The Building of the Sudbury Neutrino Observatory. Sudbury Neutrino Observatory proposal SNO-STR-87-012. *Physics in Canada* 48 (2).

Fiorentini, G., et al. 1994. *Physical Review D* 49: 6298.

Hata, N., and P. Langacker. 1994. Astrophysical Solutions Incompatible with the Solar-Neutrino Data. *Physical Review D.* 49: 3622.

Heeger, K., and R. G. H. Robertson. 1996. *Physical Review Letters* 77: 3720.

Kwong, W., and S. P. Rosen. 1995. *Physical Review D* 51: 6159.

Mikheyev, S. P., and A. Yu. Smirnov. 1985. Resonance Enhancement of Oscillations in Matter and Solar Neutrino Spectroscopy. *Soviet Journal of Nuclear Physics* 42: 1441.

Wolfenstein, L. 1979. Neutrino Oscillation in Matter. *Physical Review D* 17: 2369.

Andrew Hime is a technical staff member in the Weak Interactions Group of the Physics Division at the Los Alamos National Laboratory. He received his B.Sc. and M.Sc. in physics from the University of Guelph, Canada, in 1986 and 1988, respectively. Subsequently, Hime was awarded the distinguished overseas fellowship from the Royal Commission for the Exhibition of 1851 to pursue graduate studies at Oxford University. In 1991, based on his thesis involving precision searches for heavy neutrinos in nuclear-beta-decay spectra, Hime earned his D.Phil. degree in subatomic physics from Merton College, Oxford. Shortly thereafter, Hime joined the weak-interactions team at the Laboratory and was later nominated as an Oppenheimer Fellow. His research efforts center on fundamental investigations of the weak interaction, with emphasis upon the search for neutrino mass and the elucidation of the solar-neutrino problem. In addition, Hime has recently created a new program at the Laboratory to exploit magnetically trapped atoms for precision measurements of fundamental symmetries and, in particular, to pursue the parity-violating nature of the weak interaction. Hime is currently the associate director of the Neutral-Current Detector Program for the Sudbury Neutrino Observatory.

The Russian-American Gallium Experiment

Tom Bowles

The Soviet-American gallium experiment (SAGE) was renamed the Russian-American gallium experiment after the breakup of the Soviet Union. Los Alamos has served as the lead U.S. laboratory for the experiment, with the University of Washington, the University of Pennsylvania, Princeton University, and Louisiana State University rounding out the U.S. side of the collaboration.


Beginning life as seepage from the snow-covered slopes of Mount Elbrus, the Baksan River gradually gains momentum as it ambles slowly northwest through the rugged Caucacus Mountains. Eventually, the river rumbles past the august face of Mount Andyrchi and the incongruous cluster of buildings, homes, and shops at its base known as Neutrino Village. The Baksan Neutrino Observatory is tucked into the mountainside under about 1.6 kilometers of hard rock. About 1,000 scientists, engineers, and families are tucked into the village.

The neutrino observatory is the result of some ambitious planning on the part of Soviet scientists. In 1964, the scientists dreamed of building several large detectors dedicated to observing the evasive neutrinos that streamed unfettered through the planet. Soon, they realized that burying the experiments under tons of rock would reduce the effects of cosmic rays. The rock itself would have to be geologically stable and relatively immune to earthquakes and other natural disasters. With an eye toward saving money, the scientists thought of digging a horizontal tunnel into a steep mountain. Equipment could then be hauled around by rail rather than up and down in mineshaft elevators. The Baksan River Valley in southern Russia presented itself as the ideal site. Years later, the Institute for Nuclear Research of the Russian Academy of Sciences would build Neutrino Village for the sole purpose of accommodating the needs of the Baksan Neutrino Observatory.

The SAGE experiment is the largest research effort at Baksan. Initiated in 1985 as a collaborative effort between the United States and the former Soviet Union, the experiment was designed to measure the flux of pp neutrinos that are produced in the dominant energy-producing mechanism of the sun. That particular flux is directly tied to the measured solar luminosity and is essentially independent of solar models. Hence, observation of a significant deficit of pp neutrinos would strongly suggest that a resolution to the solar-neutrino problem lies in the properties of the neutrino, rather than in solar physics.

At present, the charge-changing interaction between electron neutrinos and a neutron in the gallium atoms provides the only feasible means to measure the low-energy pp neutrinos. The reaction transforms a stable gallium atom into a radioactive isotope of germanium:

$$\nu_e + {}^{71}\text{Ga} \rightarrow {}^{71}\text{Ge} + e^- \ .$$
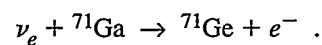
Because the unstable germanium atoms decay with a characteristic spectrum, they can be detected and their numbers counted. In this way, the solar-neutrino flux can be measured, and with a threshold of only 0.233 MeV, the reaction is sensitive to nearly the entire energy spectrum of solar neutrinos. In particular,
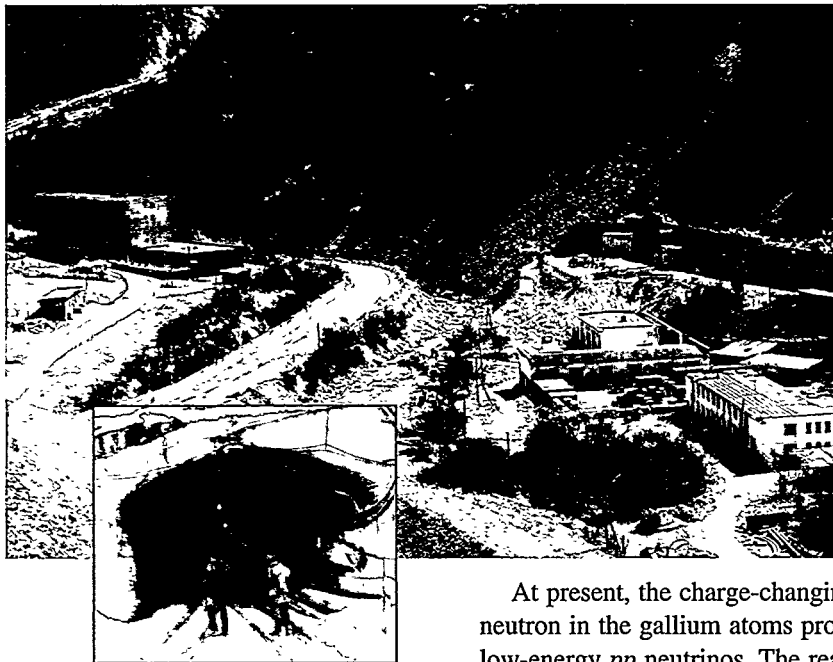
Neutrino Village lies at the base of Mount Andyrchi in the northern Caucacus Mountains. The Russians bored about 3.5 kilometers straight into the mountain in order to bury SAGE under tons of rock, and thus a researcher's access to the experiment begins with a journey through darkness.

(a)

Expose 57 tons of ^{71}Ga
metal to solar neutrinos
~30 days

Chemical extraction
of ^{71}Ge atoms
~1 day

Count number of
^{71}Ge decays
3–5 months

(b) $\quad ^{71}Ga + \nu_e \rightarrow {}^{71}Ge + e^-$

(c) $\quad ^{71}Ge + e^- \rightarrow {}^{71}Ga + \nu_e$

Energy

^{8}B

^{7}Be

pp

Solar
neutrinos

^{71}Ga

3/2$^-$

3/2$^-$

5/2$^-$

1/2$^-$

^{71}Ge*

11.4-day
half-life

^{71}Ge

Electron
capture
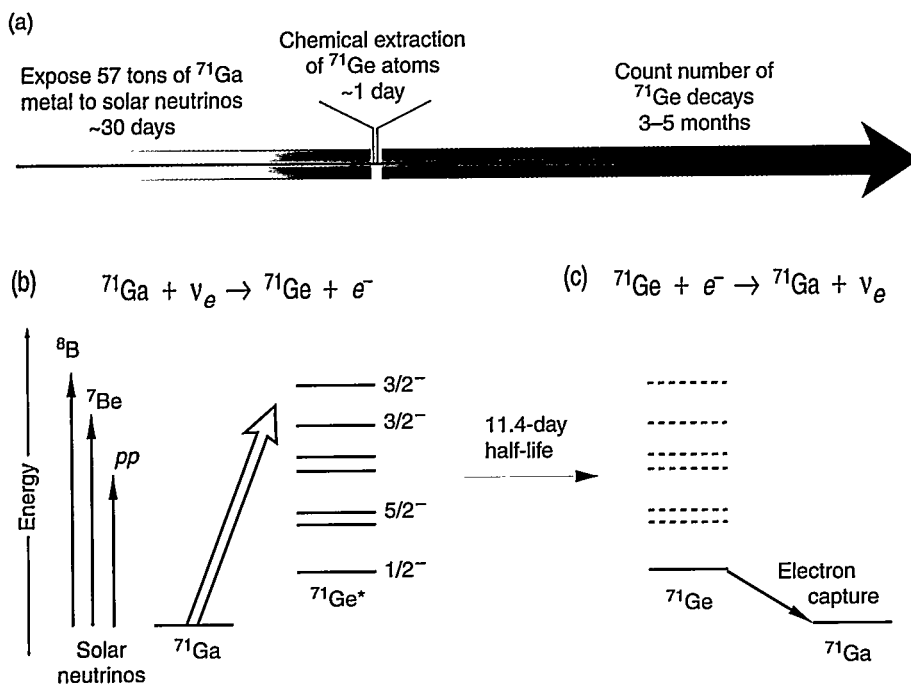
^{71}Ga

Figure 1. SAGE Overview

(a) SAGE has three distinct stages of operation: the transmutation of ^{71}Ga into unstable ^{71}Ge caused by solar neutrinos, the chemical extraction of the ^{71}Ge atoms, and the detection of the ^{71}Ge decays. The number of decays is proportional to the solar-neutrino flux.

(b) Because of low-lying excited states in the ^{71}Ge nucleus, the inverse-beta-decay reaction of gallium into germanium has a threshold of only 0.233 MeV. The reaction is therefore sensitive to pp and all other solar neutrinos. The energetic neutrinos leave the ^{71}Ge in an excited state, which quickly decays to the ground state. (c) After an 11.4-day half-life, ^{71}Ge decays by capturing an orbital electron. The remaining electrons quickly reconfigure themselves around the new nucleus and dissipate excess energy by emitting x-rays and/or Auger electrons. These emissions are the signature of the decay.

54 percent of the detected signal should be due to neutrinos from the pp reaction, based on the predictions of the standard solar model for the total solar-neutrino flux.

In addition to SAGE, GALLEX (another international collaboration headed by MPIK Heidelberg) exploits the above reaction. The composition of the gallium target differs between the two experiments. SAGE uses metallic gallium (which becomes a liquid at just above room temperature), while GALLEX uses gallium in a liquid-chloride form. The different forms of the gallium are susceptible to very different types of backgrounds, and thus the two experiments provide a check for each other. This feature helps ensure that the observed events are due to reactions of solar neutrinos on gallium, rather than some background process.

Unlike other solar-neutrino detectors, the SAGE detector has a size that is of a conceptually manageable scale. The experiment initially employed about 30 tons of liquid gallium metal distributed among 4 tanks, each the size of a small hot tub. Today, the detector contains 57 tons of liquid metal distributed among 8 tanks. The Russians provided the gallium, valued at $40 million, in addition to the chemical extraction equipment, underground laboratory, and counting and analysis facilities. The Americans provided numerous pieces of equipment, including the primary counting system. The Americans also brought to the collaboration their substantial expertise in techniques for low-level counting. The Russians were responsible for operations, but both sides participated in data collection and analysis, as well as in publication of results.

SAGE indirectly measures the solar-neutrino flux by extracting and counting the germanium atoms produced in the gallium tanks. (See Figure 1.) About once a month, a chemical extraction is performed in which individual germanium-71 (^{71}Ge) atoms are plucked from among some 5×10^{29} gallium atoms. (Only 1.2 ^{71}Ge atoms are predicted to be produced per day in 30 tons of gallium, assuming the neutrino flux predicted by the standard solar model. The efficiency of the chemical extraction is simply incredible.) Just before the extraction, about 700 micrograms of stable germanium is added to the tanks. Monitoring the recovery of this natural germanium allows a measurement of the extraction efficiency for each run. The total germanium extract is purified and synthesized into germane (GeH_4), a measured quantity of xenon is added, and the mixture is inserted into a small-volume proportional counter.

The last bit of extract containing the unstable germanium is drawn out from the gallium tank. The extract appears as a silvery pool floating on top of the duller, liquid gallium metal.

The counter is then sealed and placed in the well of a sodium iodide (NaI) detector (used as a veto), which sits inside a large, passive shield.

Germanium-71 decays with an 11.4-day half-life to gallium-71 (^{71}Ga) by electron capture, in which an orbital electron in the ^{71}Ge atom is captured by a proton in the ^{71}Ge nucleus. The nucleus is converted to ^{71}Ga, but the decay also leaves the gallium atom in an excited state. The excess energy is carried off by low-energy electrons (Auger electrons) and by x-rays produced during the electron-shell relaxation of the ^{71}Ga atom. Taken together, the electron spectrum and the x-rays make for a characteristic decay signature. Pulse-shape discrimination and a maximum-likelihood analysis identify and distinguish that signature from all the other background signals detected by the proportional counter. The number of decays occurring over 4 to 6 months is recorded.

Taking into account all efficiencies, the team expects SAGE to detect only about eight of the ^{71}Ge atoms produced in the 57 tons of gallium per run. Clearly, the backgrounds must be kept to a small fraction of a count per day. To yield such low backgrounds, the counters are made of specially selected quartz and zone-refined iron. All the components used in the NaI detector were specially selected; even the individual nuts and screws were measured for possible trace radioactivity.

The experiment began operation in May 1988 with the purification of 30 tons of gallium. Large quantities of long-lived ^{68}Ge (half-life = 271 days) had to be removed. They had been produced by cosmic rays while the gallium was on the earth's surface. By January 1990, the backgrounds had been reduced to sufficiently low levels that solar-neutrino measurements could begin.

Since that time, extraction runs have been carried out monthly except for periods dedicated to calibration runs. SAGE reports the measured value of the solar-neutrino capture rate on ^{71}Ga to be

The tops of eight greenish tanks poke through the elevated floor of the Baksan Neutrino Observatory. The squat tanks contain a total of 57 tons of ultrapure gallium. Heating the tanks to about 30 degrees centigrade keeps the gallium in liquid form. Once or twice a day, one of the hundred trillion or so solar neutrinos that pass through the tanks each second interacts with the gallium and produces an unstable germanium atom. These unstable atoms are extracted from the tanks once a month. The red motors seen perched above each tank are used to rapidly stir the liquid gallium during the extraction runs. The equipment seen in the foreground handles the chemicals used for the extraction. The red room to the left of the picture houses all the electronics and control systems used in the experiment.

$$^{71}\text{Ga capture rate} = 72 \, ^{+12}_{-10} \text{ (statistical) } ^{+5}_{-7} \text{ (systematic) SNU} \ .$$

An SNU (solar-neutrino unit) is equal to 10^{-36} captures per atom per second. This unit facilitates the comparison of results between different radiochemical experiments. The SAGE result is in excellent agreement with the GALLEX measurement of 70 ± 8 SNU. The capture rate predicted by the solar model was 132 ± 7 SNU, or nearly a factor of 2 higher.

Because SAGE observed a low signal compared with the solar-model prediction, the experiment underwent thorough checking to ensure it was working correctly. Much of the attention focused on the germanium-extraction procedure. The first test consisted of extracting stable germanium doped with a known number of radioactive ^{71}Ge atoms from 7 tons of gallium. The results indicated an extraction efficiency of 101 ± 5 percent for the natural germanium and 99 +6/ −8 percent for ^{71}Ge.

The definitive test of the extraction, however, was performed in 1995. The experiment used an extremely intense, artificial neutrino source to produce ^{71}Ge inside the detector. Chromium-51 (^{51}Cr) decays with a 27.7-day half-life by electron capture, thereby producing monoenergetic neutrinos. By placing a source containing 0.52 megacurie of ^{51}Cr inside 13 tons of gallium, one could expect to produce 50 times more ^{71}Ge atoms than the solar neutrinos would produce and anticipate to extract and observe about 147 atoms.

Half a million curies of anything is not to be treated lightly. The neutrinos

themselves are harmless, but about 10 percent of the time, ^{51}Cr decays by emitting a 320-kilo-electron-volt gamma ray. Left unshielded, those gammas would make the source a deadly menace. (Anyone holding the source, which is as small as a Coke can, would be fatally irradiated in about 1 minute.) The source was made in a fast breeder reactor in Kazakhstan. That reactor had been formerly used for plutonium production. A special assembly inserted into the reactor core was used for producing ^{51}Cr from highly enriched ^{50}Cr rods. The source was shielded by being placed inside a tungsten container with walls that were about 1 inch thick. All but about 0.001 percent of the gamma rays were absorbed in the shielding. The container radiated heat like a 100-watt light bulb.

Shielded in this manner, the source was perfectly safe to transport. But getting it into Russia turned out to be a problem. All arrangements for bringing the source in had been worked out when Kazakhstan was part of the Soviet Union, but by the time the source was to be delivered, Kazakhstan had become a separate country! The American team stood around, extremely unhappy, as 15 percent of the activity decayed away while the source sat in Russian customs for six days. A remarkable navigation of diplomatic channels finally led to a meeting, on Christmas day, between the SAGE members and the President of the Kabardino-Balkarya Academy of Sciences. Approval to transport the source to Baksan was finally granted. The result of the calibration experiment showed that the extraction and counting efficiencies were 95 ± 12 percent of those expected. The experiment was indeed working correctly. Thus, the observed deficit of solar neutrinos in SAGE is not due to some experimental error. Instead, it reflects a real deficit in neutrinos coming from the Sun.

The problem is now one of understanding the reason for the deficit. Unfortunately, SAGE cannot answer that question directly. The detection method can only infer the number of neutrino events that occurred in the gallium tanks. There is no way to extract information about the neutrino energy and, hence, about the shape of the solar-neutrino spectrum. The SAGE result, however, taken together with the results of other solar-neutrino experiments and the predictions of the standard solar model for the flux, suggests that the low-energy pp neutrinos are present in full strength while the flux of other, higher-energy solar neutrinos is significantly reduced. Many authors argue that these results cannot be reconciled with an astrophysical explanation. New experiments such as SNO are required for determining the origin of the solar-neutrino problem in a model-independent manner.

We hope that SAGE can continue measurements for a few more years. We are also investigating the possibility of converting gallium metal into gallium arsinide, which would enable construction of a real-time electronic detector for the entire solar-neutrino spectrum with very good energy resolution (a few kilo-electron-volts). The feasibility of such a detector still needs to be researched. However, the fate of the gallium, which remains the property of the Russian government, is now in doubt. Recently, an edict was issued requiring that the Institute for Nuclear Research of the Russian Academy of Sciences, which oversees the Baksan Observatory, return the gallium so that it could be sold and thus help pay salaries for unpaid government workers. While being sympathetic to the plight of the Russian workers, we also feel that, given its extremely high purity and low levels of trace radioactivity, the gallium should be considered a world resource to be invested in future research. We hope the Russian government will reach the same conclusion and explore other means to ease its fiscal crisis. ∎

Five members of the SAGE research team (the five people on the left) pose in front of the tunnel entrance. From left to right: Tanya Knodel, Ray Davis, Jr., Ken Kande, Vladimir Gavrin, and George Zatsepin. Also shown (continuing from left to right) are George Cowan, Igor Barabanov, and Keith Rowley.

A possible solution to the solar-neutrino problem

S. Peter Rosen

ver since 1968, when Ray Davis and his colleagues reported their first observations of solar neutrinos, there have been continuing reports of a deficit in the number of electron neutrinos arriving at Earth compared with the number predicted by the standard solar model. *In vacuo* neutrino oscillations, in which electron neutrinos have a certain probability of transforming into muon and/or tau neutrinos (or change flavor) as they travel between the Sun and Earth, were often viewed as one way to explain that deficit.

To date, there have been four different experiments with varying sensitivities to different parts of the solar-neutrino energy spectrum. All have reported a significant reduction in the neutrino flux. The combined data from them suggests that only neutrinos with energies between 1 and 10 million electron volts (MeV) are seriously depleted. It is unlikely that *in vacuo* oscillations alone could produce those results. Instead, the data suggest that a resonance phenomenon might be at work, one in which the probability for a neutrino to transform into another flavor is high only over a limited region of the solar-neutrino energy spectrum.

The MSW effect provides us with just such an energy-dependent phenomenon. Named after Lincoln Wolfenstein, who formulated the underlying physics, and S. P. Mikheyev and Alexei Smirnov, who recognized its importance for the Sun, the MSW effect is an elegant application of quantum mechanics that explains how neutrino oscillations can be enhanced by the medium through which the neutrinos travel. At the right density of matter, an electron

neutrino within a certain energy range can make a dramatic change to a different flavor even though its intrinsic *in vacuo* probability for doing so may be very small (Figure 1).

Neutrinos that Travel in a Vacuum

In vacuo neutrino oscillations can occur if the neutrino flavor states are "coherent," linear superpositions of neutrino mass states. Coherent means that the phases of two or more of the mass states are correlated, so that the relative phase between the states leads to an interference term in the calculation of quantum probabilities and expectation values. Quantum interference between the mass states becomes an integral part of the neutrino's description.

In certain respects, this phase phenomenon parallels the relationship between circular and plane polarization for ordinary light. The left and right states of circular, polarized light emerge most naturally from Maxwell equations. All other states can be expressed as linear superpositions of those two states. In particular, plane, polarized light is a superposition of equal amounts of two circular states that have a constant relative-phase difference. Changing the relative phase of the states rotates the plane of polarization.

Similarly, we think of the neutrino mass states $|\nu_1\rangle$, $|\nu_2\rangle$, and $|\nu_3\rangle$, with distinct masses m_1, m_2, and m_3, as the analogues of the circularly polarized states. The weak-interaction states, or the flavor neutrinos, $|\nu_e\rangle$, $|\nu_\mu\rangle$, and $|\nu_\tau\rangle$,

are created as coherent, linear superpositions of the mass states and are the analogues of the independent planes of polarization. Because the phase of each mass state $|\nu_k\rangle$ depends on the mass m_k ($k = 1, 2, 3$), each state evolves with a different phase, and therefore the relative phase between the states changes with time. Quite distinct from our analogy with ordinary light, this change can lead to the appearance of different neutrino flavors.

To keep things simple, in this article we shall consider only two neutrino mass states: $|\nu_1\rangle$ and $|\nu_2\rangle$. The electron- and muon-neutrino flavor states would then be written as

$$|\nu_e\rangle = \cos\theta|\nu_1\rangle + \sin\theta|\nu_2\rangle \text{ , and}$$
$$|\nu_\mu\rangle = -\sin\theta|\nu_1\rangle + \cos\theta|\nu_2\rangle \text{ .}$$

The angle θ characterizes the amount of mixing between the mass states and is known as the *intrinsic mixing angle*. If θ is small, the electron neutrino consists primarily of the state $|\nu_1\rangle$ and has only a small admixture of $|\nu_2\rangle$, whereas the muon neutrino would be dominated by $|\nu_2\rangle$ and would have only a small amount of $|\nu_1\rangle$.

We assume that at $t = 0$, the neutrino is created in the distinct superposition that corresponds to the electron neutrino:

$$|\nu(0)\rangle = |\nu_e\rangle = \cos\theta|\nu_1\rangle + \sin\theta|\nu_2\rangle \text{ .}$$

When the neutrino travels through a vacuum, each mass-state component $|\nu_k\rangle$ evolves with its own phase factor $\exp(-iE_k t)$. (We are working in units in which $\hbar = c = 1$.). We assume each mass state has the same momentum p, which is much greater than the masses

m_k. Thus,

$$E_k = \sqrt{p^2 + m_k^2} \approx p + \frac{m_k^2}{2p} \, ,$$

where $k = 1, 2$. Because $m_1 \neq m_2$, and hence $E_1 \neq E_2$, the relative phase between $|\nu_1\rangle$ and $|\nu_2\rangle$ will change as $|\nu(t)\rangle$ evolves with time. At an arbitrary time t, the neutrino has evolved to the state

$$|\nu(t)\rangle = \cos\,\theta\,e^{-iE_1 t}\,|\nu_1\rangle + \sin\,\theta\,e^{-iE_2 t}\,|\nu_2\rangle$$

In general, this linear combination of $|\nu_1\rangle$ and $|\nu_2\rangle$ is neither a pure electron neutrino state $|\nu_e\rangle$ nor a pure muon neutrino state $|\nu_\mu\rangle$. Instead, it is a linear superposition of both states. Quantum mechanics then tells us that, after travelling a distance of x meters, $|\nu(t)\rangle$ could be detected as a muon neutrino. That transformation probability, denoted as $P(\nu_e \rightarrow \nu_\mu)$, is given by

$$P(\nu_e \rightarrow \nu_\mu) = |\langle \nu_\mu | \nu\,(t)\rangle|^2$$

$$= \sin^2 2\theta \, \sin^2(\pi x/\lambda).$$

(Even massive neutrinos would be highly relativistic and would travel at nearly the speed of light c. We have made the approximation $t \approx x/c = x$ in our units. See the box "Derivation of Neutrino Oscillation Length" on page 161.) Because of the term $\sin^2(\pi x/\lambda)$, the probability oscillates with distance from the source. The parameter λ is called the oscillation length and is given in meters by

$$\lambda = \frac{\pi E_\nu}{1.27\Delta m^2} \, ,$$

where E_ν is the neutrino energy in million electron volts and

$$\Delta m^2 = m_2^2 - m_1^2$$

is approximately the mass difference between the muon neutrino and the electron neutrino measured in electron volts squared, assuming a small, intrinsic mixing angle. (The factor of 1.27 in the denominator allows λ to be expressed in meters. It derives in part from hidden factors of $\hbar$ and c.) Note that, if the
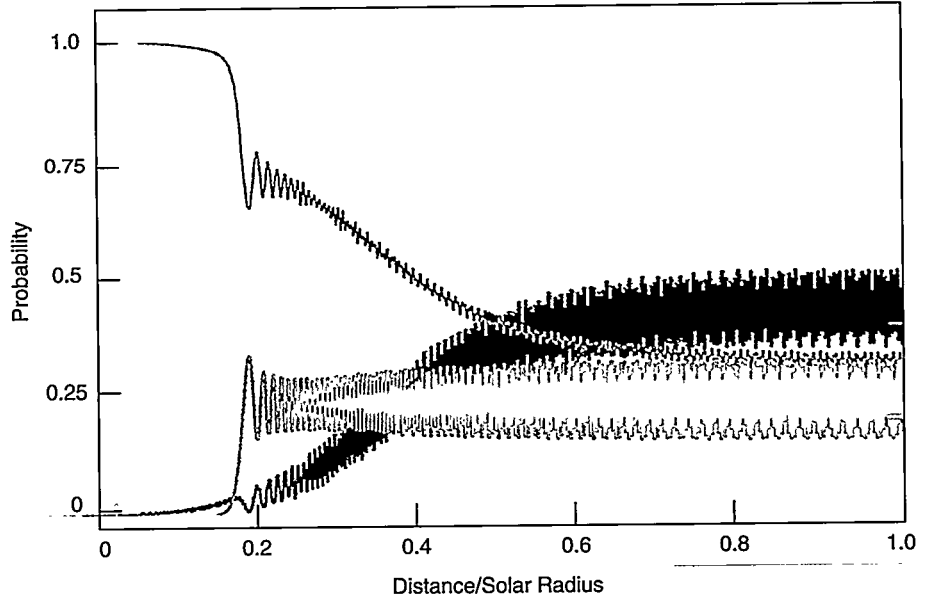

Figure 1. The Leopard Changes Its Spots

These curves represent the probability that a neutrino, after travelling through matter, would be detected as either an electron neutrino (blue), muon neutrino (red), or tau neutrino (yellow). A neutrino is "born" in the core of the sun as the superposition of mass states, in a combination that corresponds to an electron neutrino. In the specific model used to generate these curves, the superposition changes through the MSW resonance effect after the neutrino has traversed about 15 percent of the solar radius. The neutrino has only a 75 percent chance of being detected as an electron neutrino, and a 25 percent chance of being detected as a tau neutrino. By the time it flees the sun, the neutrino is most likely to be detected as a muon neutrino.

intrinsic mixing angle is small, then the oscillation probability will always be small, independent of neutrino energy.

Neutrinos that Travel through Matter

The work of Wolfenstein, and then Mikheyev and Smirnov, showed that the oscillation probability could increase dramatically because of an additional phase shift that occurs when neutrinos travel through matter. The origin of this phase shift can be understood by analogy with a well-known phenomenon: When light travels through a material, it sees a refractive index because of coherent forward scattering from the constituents of the medium. Birefringent materials have different refractive indices for independent, linear polarizations of the light, and so the

phase of each polarization component evolves differently. When polarized light passes through a birefringent material, the relative phase between the polarization states changes, and the plane of polarization rotates.

A similar phenomenon applies to the neutrino flavor states as they pass through matter. In the standard electroweak model, all neutrinos interact with up quarks, down quarks, and electrons through neutral currents (the exchange of neutral Z^0 bosons). All flavor states see a refractive index n that is a function of the neutral-current forward-scattering amplitude f_{nc}, the density of the electrons N_e, and the momentum p:

$$n_{nc} = 1 + \frac{2\pi N_e}{p^2} f_{nc} \, .$$

Electron neutrinos, and only electron neutrinos, can also interact with electrons through charged currents in a
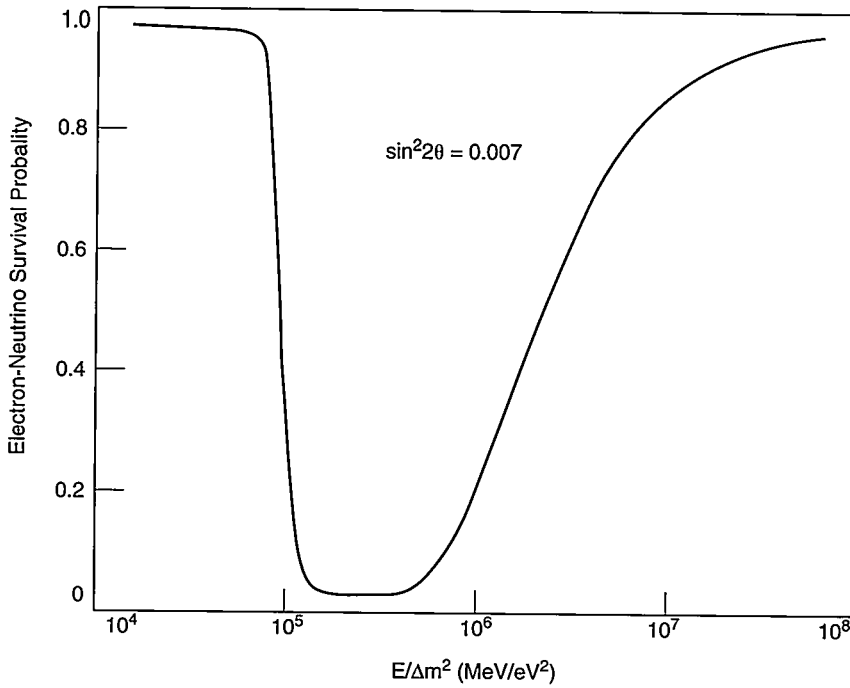
Figure 2. The MSW Survival-Probability Curve for the Sun

The probability that an electron neutrino born in the core of the Sun will emerge from the Sun as an electron neutrino is called the survival probability $P_s(\nu_e \to \nu_e)$. (The survival probability is equal to $1 - P(\nu_e \to \nu_\mu)$.) It is plotted as a function of $E_\nu/\Delta m^2$, which is essentially the *in vacuo* oscillation length λ. The mixing angle was chosen such that $\sin^2 2\theta = 0.007$, and calculation of the curve takes into account the density profile of the Sun. For a range of oscillation length values between 10^5 and 10^6 MeV/eV^2, the probability that an electron neutrino remains an electron neutrino is very small. In our two-state model, the neutrino would oscillate into a muon neutrino.

charge-changing process mediated by the W^+ boson. The refractive index seen by electron neutrinos, therefore, has an additional term n_{cc} given by

$$n_{cc} = \frac{2\pi N_e}{p^2} f_{cc} \ .$$

The charged-current forward-scattering amplitude f_{cc} in this term is proportional to the weak-interaction coupling constant G_F (the Fermi constant) times the neutrino momentum

$$f_{cc} = \frac{G_F p}{\sqrt{2}\pi} \ ,$$

so that n_{cc} takes the form

$$n_{cc} = \frac{\sqrt{2} G_F N_e}{p} \ .$$

In travelling a distance x, each flavor state develops a phase $\exp[ip(n-1)x]$ due to the index of refraction. For muon or tau neutrinos, that phase is $\exp[ip(n_{nc} - 1)x]$, while electron neutrinos have a phase given by $\exp[ip(n_{nc} + n_{cc} - 1)x]$. Substituting in the expressions for n_{nc} and n_{cc} leads to the expressions

$$e^{i2\pi x N_e f_{nc}/p} \qquad \text{for } \nu_\mu \text{ and}$$

$$e^{i(2\pi N_e f_{nc}/p + \sqrt{2} G_F N_e)x} \quad \text{for } \nu_e.$$

The additional term $\sqrt{2} G_F N_e$ in the phase of the electron neutrino is called the *matter oscillation term*. This term causes the relative phase between the electron and muon states to change with distance. Hence, the interference between the two states also changes

with distance and, as always, is accompanied by interference phenomena. The interference can be totally destructive, totally constructive, or somewhere in between, depending upon relative conditions.

The equation for the MSW probability is derived in a heuristic fashion in the box on page 161. Here, we simply state the results. The MSW probability for an electron neutrino to transform into a muon neutrino is

$$P_{MSW}(\nu_e \to \nu_\mu) = \sin^2 2\theta_m \sin^2\left(\frac{\pi x W}{\lambda}\right) ,$$

where

$$\sin^2 2\theta_m = \frac{\sin^2 2\theta}{W^2} \ ,$$

$$W^2 = \sin^2 2\theta + (D - \cos 2\theta)^2 \text{ , and}$$

$$D = \sqrt{2} G_F N_e \frac{2E_\nu}{\Delta m^2} \ .$$

Note the similarity between this expression for the MSW probability and the *in vacuo* oscillation probability. The *in vacuo* mixing angle θ is replaced by an effective mixing angle θ_m that depends on the matter oscillation term (through the parameter D). When a neutrino travels in vacuum, the electron density is zero, and hence D is equal to zero, so that $W^2 = 1$. Thus, the MSW probability reduces to the *in vacuo* probability.

When a neutrino travels through matter, however, W^2 can become less than 1. This is the MSW resonance effect. The oscillation probability increases with the resonance condition given by $D = \cos 2\theta$. At that point, $W^2 = \sin^2 2\theta$ and $\sin^2 2\theta_m = 1$, and the oscillation probability reaches a maximum. This resonance condition is independent of the size of the intrinsic mixing angle θ, but it requires matching the properties of the material with the neutrino oscillation length λ through the relation

$$\sqrt{2} G_F N_e = 1.2 \frac{\cos 2\theta}{\lambda} \ .$$

When this matching occurs, a neutrino will have a high probability

of transforming into a neutrino of a different flavor, even if its intrinsic *in vacuo* probability for doing so might be very small. This enhancement is called the MSW effect.

Oscillation Enhancement in the Sun

The density of the Sun is not constant but decreases monotonically from about 150 grams per cubic centimeter (g/cm³) at the center to 0.1 g/cm³ at a radius of 700,000 kilometers. In other words, the density decreases from 50 times the density of terrestrial rocks to one-tenth the density of water. For this range of densities, the value of the parameter D varies from a few tens near the core to a few hundredths near the edge. It passes through 1 at some intermediate point. Large values of D lead to very small values for the effective mixing angle and damp out neutrino oscillations, whereas small values essentially leave the *in vacuo* oscillation unchanged. Values near 1 give rise to the maximum enhancement (we are assuming that θ is small).

An electron neutrino born in the core of the Sun will start its journey to the edge without oscillating—until it reaches the region where D is of the order of 1. In this region, it will undergo the oscillation enhancement, and the probability that it will remain an electron neutrino will decrease significantly. In Figure 1, this region is observed at a distance of 0.2 solar radius. At larger solar radii, D is much smaller, and the electron neutrino's survival probability will oscillate around a relatively small value with an amplitude corresponding to the *in vacuo* mixing angle.

We can estimate the range of λ or Δm^2 for which the MSW effect is important. We rewrite the enhancement condition in terms of the solar density ρ_e measured in grams per cubic centimeter and then multiply by Avogadro's number. For the neutrino energy measured in million electron volts and Δm^2 measured in electron
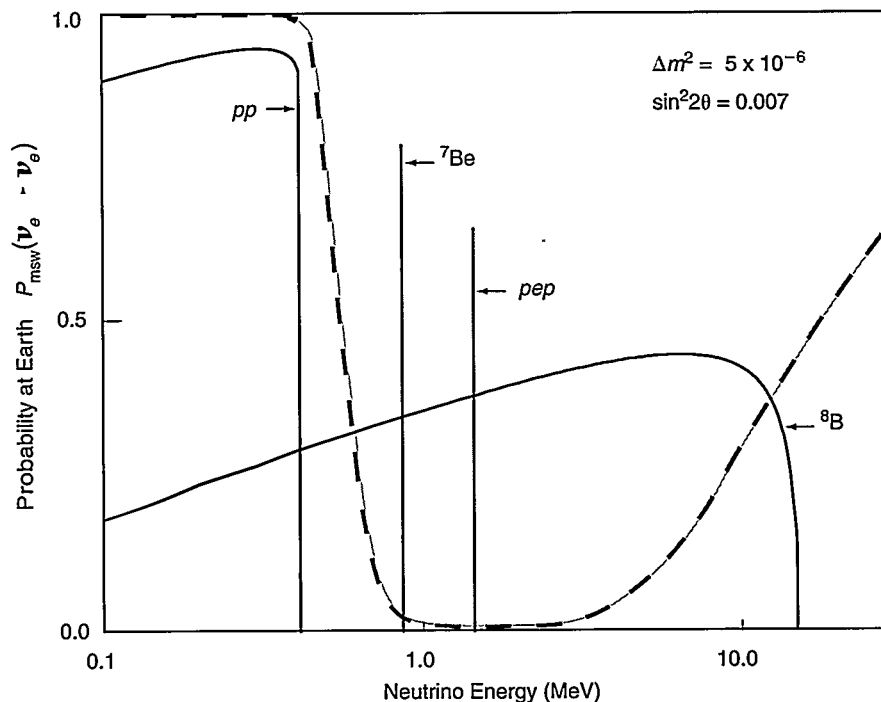

Figure 3. MSW Solution to the Solar-Neutrino Problem

The MSW survival probability $P_{MSW}(\nu_e \to \nu_e)$ (red dashed line) has been plotted as a function of neutrino energy and superimposed over a simplified picture of the solar-neutrino spectrum. Solar neutrinos with energies that fall within the gray-shaded region (those from the ^{7}Be, *pep*, and ^{8}B reactions) have a low survival probability, and thus have a high probability of oscillating into muon and/or tau neutrinos as they flee the solar core. If the MSW effect occurs in the Sun, experiments that detect only electron neutrinos would measure a reduced flux from those reactions. An experiment such as the one planned at the Sudbury Neutrino Observatory, which is sensitive to electron, muon, and tau flavors, should detect the full flux.

volts squared (eV²),

$$2.5\,\lambda = \frac{7 \times 10^6}{\rho_e}\cos 2\theta \ .$$

The *in vacuo* oscillation length λ is measured in meters for a neutrino with given momentum and mass parameters. The scale of the right-hand side is 1.8×10^4 kilometers, which is comparable with Earth's diameter. Given the range of solar densities, the MSW effect could occur if λ were in the range

$$10^4 \le \lambda \le 10^8 \text{ meters} \ ,$$

as seen in Figure 2.

Because $\lambda = \pi E_\nu/1.27\Delta m^2$ and because solar-neutrino energies vary from a fraction of 1 MeV to about 10 MeV, the MSW effect can occur in the Sun if the squared

mass differences are

$$10^{-4} \ge \Delta m^2 \ge 10^{-9}\,\text{eV}^2 \ .$$

To study this range of mass differences with terrestrial neutrinos, we would need intense sources of extremely low energy neutrinos, well below 1 MeV. Those sources do not exist, and therefore solar neutrinos are the only means available to us.

The Most Favored Solution to the Solar-Neutrino Problem

The value of Δm^2 determines how the curve for the survival probability overlays the spectrum of solar neutrinos. Decreasing Δm^2 moves the curve to the left, while increasing it moves

the curve to the right. For values of roughly $\Delta m^2 = 5 \times 10^{-6}$ eV^2 and $\sin^2 2\theta = 0.007$, we find that pp neutrinos survive as electron neutrinos most of the time, while ^{7}Be neutrinos are almost completely converted to muon neutrinos (Figure 3). This appears to be a good description of data measured by the current generation of solar-neutrino experiments. (See the article "Exorcising Ghosts" on page 136.) Next-generation experiments, such as the one planned at the Sudbury Neutrino Observatory, are designed to determine whether oscillations to other neutrino states do indeed occur and whether the MSW effect or *in vacuo* oscillations solve the solar-neutrino problem.

There is one final, but very interesting, comment. Our planet Earth may play a unique role in the study of the MSW effect. There turns out to be a well-defined range of mixing angles and mass differences for which the enhancement density is less than 15 g/cm^3. This density occurs in both the Sun and Earth, and thus neutrinos that are converted from electron neutrinos to muon neutrinos in the Sun may be reconverted to electron neutrinos when they pass through Earth.

This effect would be seen as a significant increase in the solar-neutrino signal at night, when Earth is between the Sun and the detector. Observation of such a "day-night" effect would be an unambiguous and definitive proof of the MSW effect and of neutrino mass. It would also be nature's tip-of-the-hat to the insightful Lincoln Wolfenstein, who once observed that "...for neutrinos, the Sun shines at night!" ■

S. Peter Rosen, a native of London, earned his B.S. in mathematics and an M.A. and Ph.D. in theoretical physics from Merton College, Oxford, in 1954 and 1957, respectively. Rosen came to the United States as a research associate at Washington University, and from 1959 through 1961 he worked as scientist for the Midwestern Universities Research Association at Madison, Wisconsin. In 1961, he was awarded a NATO Fellowship to the Clarendon Laboratory at Oxford. Rosen returned to the United States as a Professor of Physics at Purdue University and later served as senior theoretical physicist for the High Energy Physics Program of the U.S. Energy Research and Development Administration. Some of Rosen's additional appointments include program associate for theoretical physics with the National Science Foundation, chairman of the U.S. Department of Energy's Technical Assessment Panel for Proton Decay and of Universities Research Association, Inc., Board of Overseers at Fermi National Accelerator Laboratory. Rosen first came to Los Alamos in 1977 as a visiting staff member and later became associate division leader for Nuclear and Particle Physics of the Theoretical Division. In 1990, he accepted the position of dean of science and professor of physics at the University of Texas, Arlington, and remained in that position until January 1997, when he was appointed associate director for High Energy and Nuclear Physics at the Department of Energy in Washington, D.C. He is a fellow of the American Physical Society and of the American Association for the Advancement of Science.

Further Reading

Rosen, S. P., and J. M. Gelb. 1986. Mikheyev-Smirnov-Wolfenstein Enhancement of Oscillations as a Possible Solution to the Solar-Neutrino Problem. *Physical Review D* **34** (4): 969.

Wolfenstein, L. 1978. Neutrino Scintillations in Matter. *Physical Review D* **17** (9): 2369.

Heuristic Derivation of the MSW Effect
(for the students in us all)

For simplicity, we shall consider only oscillations between electron and muon neutrinos. The neutrino mass states $|\nu_1\rangle$ and $|\nu_2\rangle$ are assumed to have distinct masses m_1 and m_2, respectively. We define the neutrino flavor states $|\nu_e\rangle$ and $|\nu_\mu\rangle$ in terms of two mass states:

$$|\nu_e\rangle = \cos\theta\,|\nu_1\rangle + \sin\theta\,|\nu_2\rangle \ , \text{ and} \qquad (1a)$$

$$|\nu_\mu\rangle = -\sin\theta\,|\nu_1\rangle + \cos\theta\,|\nu_2\rangle \ . \qquad (1b)$$

We further assume that an electron neutrino is born at time $t = 0$. That neutrino will evolve in time as a superposition of states with time-dependent coefficients. The neutrino can be described by either mass states or flavor states:

$$|\nu(t)\rangle = a_1(t)\,|\nu_1\rangle + a_2(t)\,|\nu_2\rangle = a_e(t)\,|\nu_e\rangle + a_\mu(t)\,|\nu_\mu\rangle \ , \qquad (2)$$

where

$$a_e(t) = a_1(t)\cos\theta + a_2(t)\sin\theta \ , \text{ and} \qquad (3a)$$

$$a_\mu(t) = -a_1(t)\sin\theta + a_2(t)\cos\theta \ . \qquad (3b)$$

We will also be using the inverse of Equations (1) and (3):

$$|\nu_1\rangle = \cos\theta\,|\nu_e\rangle - \sin\theta\,|\nu_\mu\rangle \ , \text{ and} \qquad (4a)$$

$$|\nu_2\rangle = \sin\theta\,|\nu_e\rangle + \cos\theta\,|\nu_\mu\rangle \ ; \qquad (4b)$$

$$a_1(t) = a_e(t)\cos\theta - a_\mu(t)\sin\theta \ , \text{ and} \qquad (4c)$$

$$a_2(t) = a_e(t)\sin\theta + a_\mu(t)\cos\theta \ . \qquad (4d)$$

In general, the time development of the neutrino states described in Equation (2) has a phase that depends on both the momentum and the energy of the neutrino. For example, an electron neutrino evolves as

$$|\nu_e(t)\rangle = \cos\theta\, e^{ip\cdot x - iE_1 t}\,|\nu_1\rangle + \sin\theta\, e^{ip\cdot x - iE_2 t}\,|\nu_2\rangle \ . \qquad (5)$$

We work in units in which $\hbar = c = 1$. Let us first consider the evolution of $|\nu(t)\rangle$ as a superposition of mass eigenstates during an infinitesimal time Δt. We assume a common momentum for each mass state, so that only the difference between the energies of the mass states (due to the difference in the neutrino masses) governs the time development of the state. With $p \gg m_k$, we can approximate the energy as

$$E_k = \sqrt{p^2 + m_k^2} \ \approx \ p + m_k^2/2p = p + M_k \ , \qquad (6)$$

where $M_k = m_k^2/2p$ ($k = 1, 2$). The neutrino evolves in time Δt as

$$|\nu(t+\Delta t)\rangle = a_1(t+\Delta t)e^{-iE_1\,\Delta t}\,|\nu_1\rangle + a_2(t+\Delta t)e^{-iE_2\,\Delta t}\,|\nu_2\rangle$$

$$\approx a_1(t+\Delta t)\, e^{-iM_1\,\Delta t}\,|\nu_1\rangle + a_2(t+\Delta t)e^{-iM_2\,\Delta t}\,|\nu_2\rangle \ . \qquad (7)$$

We have dropped the overall phase factor of $\exp(-ip\Delta t)$ in Equation (7) because it has no bearing on the final result. With the help of Equations (4a) and (4b), we can write Equation (7) in the flavor basis:

$$|\nu(t+\Delta t)\rangle = [a_1(t+\Delta t)e^{-iM_1\,\Delta t}\cos\theta + a_2(t+\Delta t)e^{-iM_2\,\Delta t}\sin\theta]\,|\nu_e\rangle$$
$$+ [-a_1(t+\Delta t)e^{-iM_1\,\Delta t}\sin\theta + a_2(t+\Delta t)e^{-iM_2\,\Delta t}\cos\theta]\,|\nu_\mu\rangle \ . \tag{8}$$

We next consider the neutrino as a superposition of flavor states:

$$|\nu(t)\rangle = a_e(t)\,|\nu_e\rangle + a_\mu(t)\,|\nu_\mu\rangle \ . \tag{9}$$

Because only electron neutrinos interact via charged currents, the two flavor states have different forward-scattering amplitudes, and each sees a different effective refractive index in matter. We assume that the change in the probability amplitudes $a_e(t)$ and $a_\mu(t)$ during an infinitesimal time Δt can be expressed as a simple phase shift that is proportional to the refractive index:

$$a_e(t+\Delta t) \approx a_e(t)\exp[ip\,(n_{nc}+n_{cc}-1)\Delta t] = a_e(t)\exp[i(\xi+\eta)\Delta t \ , \text{ and} \tag{10a}$$

$$a_\mu(t+\Delta t) \approx a_\mu(t)\exp[ip\,(n_{nc}-1)\Delta t] = a_\mu(t)\exp[i\xi\,\Delta t] \ , \tag{10b}$$

where $\xi = (2\pi N_e/p)f_{nc}$ and $\eta = \sqrt{2}G_F N_e$. The latter relation is the matter oscillation term. We have also used $\Delta x \approx \Delta t$. The neutrino state, therefore, evolves as

$$|\nu(t+\Delta t)\rangle = a_e(t)e^{i(\xi+\eta)\Delta t}\,|\nu_e\rangle + a_\mu(t)e^{i\xi\,\Delta t}\,|\nu_\mu\rangle$$
$$= a_e(t)e^{i\eta\Delta t}\,|\nu_e\rangle + a_\mu(t)\,|\nu_\mu\rangle \ , \tag{11}$$

where again we have dropped the overall phase factor of $\exp(i\xi\Delta t)$ because it does not affect the final result. Equations (8) and (11) are expressions for $|\nu(t+\Delta t)\rangle$. Equating the coefficients of $|\nu_e\rangle$ and $|\nu_\mu\rangle$ results in a set of coupled equations:

$$a_1(t+\Delta t)e^{-iM_1\,\Delta t}\cos\theta + a_2(t+\Delta t)e^{-iM_2\,\Delta t}\sin\theta = a_e(t)e^{i(\xi+\eta)\Delta t} \ , \tag{12a}$$
$$-a_1(t+\Delta t)e^{-iM_1\,\Delta t}\sin\theta + a_2(t+\Delta t)e^{-iM_2\,\Delta t}\cos\theta = a_\mu(t)e^{i\xi\,\Delta t} \ . \tag{12b}$$

Both sides of Equations (12a) and (12b) are expanded to first order in Δt,

$$[a_1(t) + \dot{a}_1(t)\Delta t - ia_1(t)M_1\Delta t]\cos\theta + [a_2(t) + \dot{a}_2(t)\Delta t - ia_2(t)M_2\Delta t]\sin\theta = a_e(t)(1+i\eta\Delta t) \tag{13a}$$

$$-[a_1(t) + \dot{a}_1(t)\Delta t - ia_1(t)M_1\Delta t]\sin\theta + [a_2(t) + \dot{a}_2(t)\Delta t - ia_2(t)M_2\Delta t]\cos\theta = a_\mu(t) \ , \tag{13b}$$

where a dot indicates the time derivative. Equations (4c) and (4d) are used to express $a_1(t), \dot{a}_1(t), a_2(t),$ and $\dot{a}_2(t)$ in terms of $a_e(t), \dot{a}_e(t), a_\mu(t),$ and $\dot{a}_\mu(t)$. Following more algebraic operations,

$$-i\dot{a}_e(t) = [M_1\cos^2\theta + M_2\sin^2\theta + \eta]a_e(t) + (M_2-M_1)\cos\theta\sin\theta\,a_\mu(t) \ , \tag{14a}$$

$$-\dot{a}_\mu(t) = (M_2-M_1)\cos\theta\sin\theta\,a_e(t) + [M_1\sin^2\theta + M_2\cos^2\theta]a_\mu(t) \ . \tag{14b}$$

These expressions can be cast in a Schrödinger-like equation for a column matrix A consisting of the probability amplitudes $a_e(t)$ and $a_\mu(t)$:

$$-i\frac{dA}{dt} = HA \ , \tag{15}$$

where $H = \begin{pmatrix} M_1\cos^2\theta + M_2\sin^2\theta + \eta & (M_2-M_1)\cos\theta\sin\theta \\ (M_2-M_1)\cos\theta\sin\theta & M_1\sin^2\theta + M_2\cos^2\theta \end{pmatrix}$ and $A = \begin{pmatrix} a_e(t) \\ a_\mu(t) \end{pmatrix}$

The eigenvalues of the matrix H are given by

$$\chi_{1,2} = \frac{\eta + M_1 + M_2}{2} \mp \frac{\sqrt{\eta^2 + (M_2 - M_1)^2 - 2\eta(M_2 - M_1)\cos 2\theta}}{2} \quad . \tag{16}$$

Equation (15) can then be solved:

$$A(t) = \left[\frac{\chi_2 e^{i\chi_1 t} - \chi_1 e^{i\chi_2 t}}{\chi_2 - \chi_1} I + \frac{e^{i\chi_2 t} - e^{i\chi_1 t}}{\chi_2 - \chi_1} H \right] A(0) \quad , \tag{17}$$

where I is the identity matrix. At time $t = 0$, the beam consists only of electron neutrinos. Thus, $a_e(0) = 1$, and $a_\mu(0) = 0$ so that

$$a_e(t) = \frac{\chi_2 e^{i\chi_1 t} - \chi_1 e^{i\chi_2 t}}{\chi_2 - \chi_1} + \frac{e^{i\chi_2 t} - e^{i\chi_1 t}}{\chi_2 - \chi_1}(M_1\cos^2\theta + M_2\sin^2\theta + \eta) \quad ,\tag{18a}$$

$$a_\mu(t) = \frac{e^{i\chi_2 t} - e^{i\chi_1 t}}{\chi_2 - \chi_1}(M_2 - M_1)\cos\theta \sin\theta \quad . \tag{18b}$$

The probability of detecting a muon neutrino after a time t is given by

$$P_{\text{MSW}}(\nu_e \to \nu_\mu) = |\langle\nu_\mu|\nu(t)\rangle|^2 = |a_e(t)\langle\nu_\mu|\nu_e\rangle + a_\mu(t)\langle\nu_\mu|\nu_\mu\rangle|^2 = |a_\mu(t)|^2 \tag{19}$$

so that

$$P_{\text{MSW}}(\nu_e \to \nu_\mu) = \frac{(M_2 - M_1)^2 \cos^2\theta \sin^2\theta}{(\chi_2 - \chi_1)^2} 2(1 - \cos(\chi_2 - \chi_1)t)$$

$$= \frac{(M_2 - M_1)^2 \sin^2 2\theta}{(\chi_2 - \chi_1)^2} \sin^2\frac{(\chi_2 - \chi_1)}{2} t \quad . \tag{20}$$

By substituting in the expressions for χ_1, χ_2, M_1, M_2, we have

$$(\chi_2 - \chi_1)^2 = (M_2 - M_1)^2 \left[\sin^2 2\theta + \left(\frac{\eta}{M_2 - M_1} - \cos 2\theta \right)^2 \right] \tag{21a}$$

$$(M_2 - M_1) = \frac{\Delta m^2}{2p} \approx \frac{\Delta m^2}{2E_\nu} \quad . \tag{21b}$$

Recalling that $x = t$, and $\eta = \sqrt{2}G_F N_e$, we arrive at the MSW probability for an electron neutrino to oscillate into a muon neutrino:

$$P_{\text{MSW}}(\nu_e \to \nu_\mu) = \frac{\sin^2 2\theta}{W^2} \sin^2\left(\frac{\pi x W}{\lambda} \right) \quad , \tag{22}$$

$$W^2 = \sin^2 2\theta + \left(\sqrt{2}G_F N_e \frac{2E_\nu}{\Delta m^2} - \cos 2\theta \right)^2 \quad , \tag{23}$$

where λ is the *in vacuo* oscillation length,

$$\lambda = 2\pi \frac{2E_\nu}{\Delta m^2} \quad , \tag{24}$$

and $\Delta m^2 = m_2^2 - m_1^2$ is required to be nonzero. If the numerical values of the hidden factors of $\hbar$ and c are included, the expression for the oscillation length becomes $\lambda = \pi E/1.27\Delta m^2$. ∎

Neutrinos and Supernovae

Marc Herant, Stirling A. Colgate, Willy Benz, and Chris Fryer

Evanescent, fleeting, transient—these words come to mind when describing the elusive neutrino. Although neutrinos clearly play a key role in nuclear physics through the weak force, their interactions with matter are just that—weak. Under typical conditions, a neutrino is 100 billion billion (10^{20}) times less likely than light to interact with matter, and a neutrino will pass straight through our planet Earth as effortlessly as the breeze through an open window. Is it any wonder that the direct physical manifestations of the neutrino always seem so tenuous?

But there is one exception to the neutrino's demure role. It occurs in the heart of massive stars, deep within the stellar core. When a massive star dies, it does not go peacefully. Instead, it makes a spectacular exit—the most powerful explosion known to occur in the universe. Astrophysicists call this exploding star a supernova, and in an ironic reversal of roles, it is the quiet neutrino that is chiefly responsible for the cataclysm.[1]

Over the years, scores of researchers (including quite a few from Los Alamos who have a particular interest in large explosions) have constructed an in-depth theory explaining how and why massive stars explode. Stars emit light and shine because they "burn" nuclear fuel. But the amount of nuclear fuel is limited.

When a star exhausts its fuel supply, something startling happens: the forces that support the star's core quickly retreat, and the core is almost instantly crushed by gravity. The compression is

so severe that, in less than 1 second, the core reaches virtually unparalleled conditions of temperature and density. Theoretical physics predicts that, under these unique circumstances, vast quantities of neutrinos are produced that carry off the enormous amount of energy
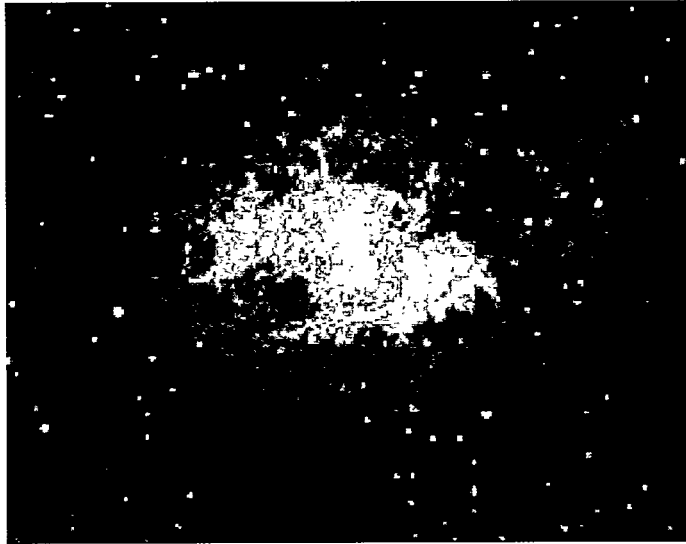

Figure 1. The Crab Nebula
Located about 6,500 light years from Earth, the crab nebula is an expanding gaseous cloud that was hurled into space when a giant star exploded. That supernova was visible day and night for several weeks in July 1054. Even today, the visible emission from the nebula is still greater than 75,000 suns. At the center of this brilliantly glowing cloud lies a spinning neutron star, which is the core of the original star.
© Malin/Pasachoff/Caltech, color photograph by David Malin.

released by the collapse of the core. A few of those neutrinos are absorbed by material that is plummeting toward the compacted core. The falling matter becomes very hot, expands, and surges outward. Eventually, the star erupts in a furious explosion that ejects the star's outer layers into space. All that remains of the once enormous star is its center, now transformed into a tiny, incredibly dense object called a neutron star.

This pivotal and wondrous function of the neutrino, so much in contrast with its usual marginal position, received triumphant vindication in February 1987, when two underground detectors recorded a burst of neutrinos and a spectacular supernova was later

observed by astronomers worldwide. The astrophysical community was elated! For the first time, the theoretical relationship between neutrinos and supernovae was empirically confirmed.

That confirmation was a climactic moment in a long history of supernovae observations. For centuries, mankind has been fascinated by the sudden, yet brief, appearance in the sky of a superbright star at a spot where there was none before. Chinese astronomers recorded one such event as early as 185 A.D. But such sightings are rare, as supernovae are infrequent events. They occur on average only once every 50 years or so within a given galaxy. The inhabitants of the northern hemisphere have not been treated with a supernova visible to the naked eye since 1604.

But it is also true that there are billions of distant galaxies within the universe, and supernovae tend to be highly conspicuous. So much energy is released during the explosion that, for a short time, the star may outshine an entire galaxy containing over ten billion stars. In the last hundred years, astronomers have monitored more than a thousand supernovae. They have been able to examine in great detail the expanding nebulae that linger for centuries as remnants of the explosions (Figure 1).

Indeed, astrophysicists have been able to study even the exotic neutron stars that form under the remarkable conditions found inside supernovae. Neutron stars are made up almost ßentirely of neutrons. Only 20 kilometers or so in diameter but more massive than the Sun, these singular objects are so dense that a basketball-sized chunk would weigh about 10 trillion tons. Their possible existence was predicted

[1]Supernovae are classified as Type I or II. Type I supernovae have no hydrogen in their emission spectra and are generated (usually) by old stars of small mass. In this article, we will only describe the more common, Type II supernovae.

by J. Robert Oppenheimer and George Volkoff in the late 1930s, and then Fritz Zwicky suggested that neutron stars might be created in supernova explosions.

Neutron stars remained but a theoretical conjecture until Jocelyn Bell and others discovered pulsars in 1967. Pulsars are often found at the center of supernova nebulae. They emit extremely regular, very intense pulses of radio waves. Only a spinning star with a diameter comparable to the breadth of a small city could lead to such an extraordinary extraterrestrial signal, and pulsars were quickly identified with neutron stars.

In this article, we outline much of what has been learned about Type II supernovae and describe in detail how old stars of more than 8 solar masses are thought to die. (A star's mass is always stated relative to the Sun's mass, which is 2×10^{33} grams and is denoted by the symbol $M_\odot$. Therefore, 8 solar masses is written as $8M_\odot$.) However, before we discuss the death of stars, we will digress and first discuss how those stars live.

A Star's Life

A star performs one of nature's finest high-wire acts. It carefully and continuously maintains its balance against the omnipresent pull of gravity. It is gravity that initially shapes a primordial cloud of gas[2] into a spherical star, and it is gravity that collapses and compresses the gas. Compression, however, increases both the temperature and the internal pressure of the gas. Once that pressure is sufficient to counteract gravity's pull, the star stops shrinking. If for some reason the internal pressure temporarily exceeds the gravitational force, the star will expand.

The pressure will then drop, and the expansion will stop once the pressure is again equal to gravity. As long as the internal pressure can be sustained, a star will neither expand nor contract, but it will maintain a state of *hydrostatic* equilibrium, wherein gravity and the internal pressure are balanced.

But a star is also hot, with a core temperature of millions of kelvins. Heat and energy flow out from the core and through the mantle to be emitted as light from the star's surface. The star shines brilliantly. Yet for all its serene beauty, starlight is a relentless drain on the star because energy is irretrievably lost to the cold vacuum of space. If energy were not continually regenerated, the loss would cool the gas and sap the internal pressure, causing the star to slowly contract.

New energy comes from thermonuclear fusion, the process whereby two light, atomic nuclei merge to form a single, heavier nucleus. Because fusion releases a significant amount of energy, the star can counteract radiative losses simply by sustaining a sufficient fusion rate. A star achieves and maintains a *thermal* equilibrium in addition to its hydrostatic equipoise. A star's life consists of balancing the opposing forces of gravity and pressure, while simultaneously matching all energy losses with the gains produced by thermonuclear fusion.

Evidently, this state of total equilibrium cannot be maintained. The amount of nuclear fuel available to the star is finite, and as lighter elements burn, fuel slowly disappears. Initially, it is only the primordial hydrogen that burns. The burning takes place in the core, which is the hottest and densest part of the star. (See the article "Exorcising Ghosts" on page 136 for a description of the energy-producing reactions in the Sun.) In part because hydrogen burning releases a lot of energy, only a modest rate of fusion is needed to stabilize the star, and the hydrogen reserves last a long time. A star will burn hydrogen for millions to trillions of years.[3]

At some point, however, all the hydrogen in the core will have fused into helium. Because helium burning requires much higher core temperatures and densities than exist at this stage of the star's life, fusion temporarily stops. Without an energy source, the core begins to cool, the core pressure begins to drop, and gravity again compresses the star. As before, the gravitational compression does work on the stellar gas so that, somewhat counterintuitively, the loss of fusion energy leads to a *rise* in the core temperature. Once the temperature and density are sufficient to fuse helium into carbon, new energy is released, and equilibrium is quickly restored. The star still consists almost entirely of hydrogen gas, but the hydrogen now surrounds a helium gas core that is undergoing fusion.

Eventually, the helium fuel is depleted. Fusion stops, and the star cools and contracts until it is again able to initiate the burning of a new fuel. This is a repetitive process, so that the aging star will burn in succession carbon, neon, oxygen, and finally silicon. Because of the various burning stages, the star develops a layered structure consisting of many different elements, as seen in Figure 2.

However, as the elements get heavier, the amount of energy released per reaction decreases. As a result, the burning rate must increase in order to liberate enough energy to sustain the internal core pressure. In addition, neutrinos are produced much more readily within the core during the late burning stages of stellar evolution. Because the neutrinos remove even more energy from the core, they are yet another factor that leads to an increased burning rate. (See the box "The Urca Process" on page 168.)

---

[2]The primordial gas consists of hydrogen, some helium, and trace amounts of other light elements. This gas formed in the first few minutes after the Big Bang. See the article "Dark Matter and Massive Neutrinos" on page 180 for more details.

[3]The time it takes for a star to burn its fuel decreases rapidly as a star's mass increases. Compared with their lighter cousins, massive stars are squeezed harder by gravity and therefore require significantly more pressure to remain stable. They burn their fuel considerably faster. Whereas the Sun will live approximately 20 billion years, a 15 $M_\odot$ star will only live about 20 million years.

**Primordial
gas cloud**
77% hydrogen
23% helium
(by mass)

Hydrogen burning
Core temperature = 4×10^7 K
Core density = 7 g/cm^3

Helium burning
Core temperature = 1.6×10^8 K
Core density = 1,500 g/cm^3

10 to 15 million years

1 million years

End of silicon burning
(Enlarged view)

Hydrogen ≈ 23,000,000 km

**Carbon
burning**
100,000 years

Helium ≈ 500,000 km

Neon burning
<1 year

Carbon, oxygen, neon ≈ 36,000 km

Oxygen burning
10,000 years

The core
Silicon ≈ 4,000 km
Iron ≈ 850 km
Central temperature > 10^9 K
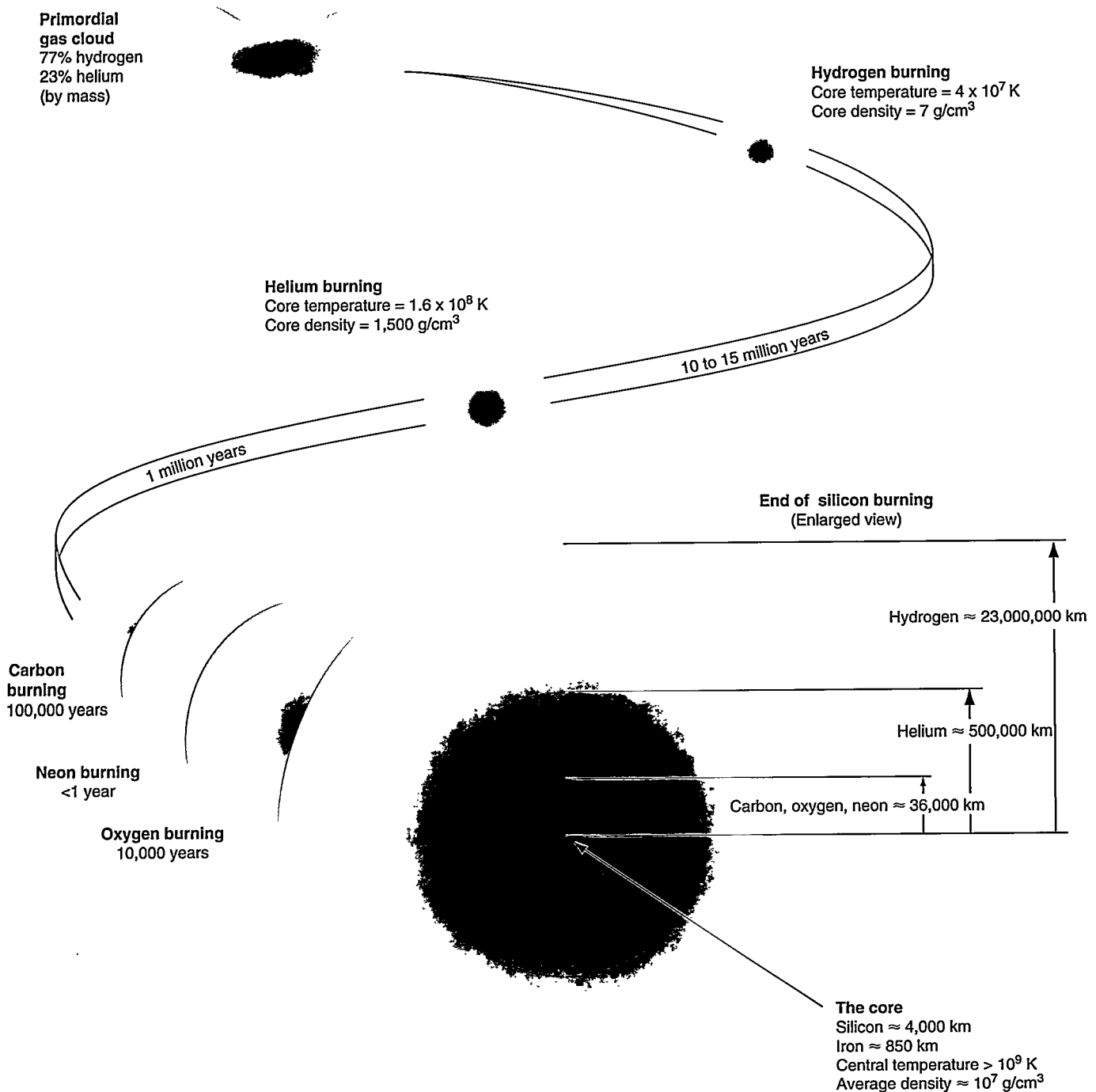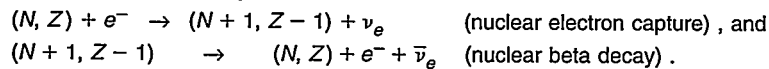Average density ≈ 10^7 g/cm^3

Figure 2. The Life of a Massive Star

A star is born when a huge cloud of primordial gas is compressed by gravity. The compression raises the density and temperature of the gas to the point that hydrogen nuclei can fuse into helium within the star's core. Both hydrostatic and thermal equilibria are quickly established (see text). The star will burn hydrogen for tens of millions of years, gradually accumulating helium in its core. Eventually, the core is fully depleted of hydrogen, and fusion stops. The core cools and contracts, which leads to higher pressures and densities, and a new burning phase begins. Helium is fused into carbon within a hotter, denser, and much smaller core, even though the star itself has become larger during this phase. Over the course of its lifetime, the star's core will become smaller and much denser as it burns in succession carbon, neon, oxygen, and silicon. At the end of silicon burning, the star has developed a layered structure, shown above for an $18M_\odot$ star. Note the tiny silicon and iron core. The core is 100 million times more dense than the hydrogen layer.

The Urca Process

During the late burning stages, core electrons become energetic enough to react with protons inside heavy nuclei through the weak interaction. The proton turns into a neutron while the electron (e^-) turns into an electron neutrino (ν_e). The neutrino escapes from the core and removes energy. The newly formed nucleus then undergoes beta decay. As a result, the nucleus is restored to its original state, and an electron–electron antineutrino ($\bar{\nu}_e$) pair is created. The $\bar{\nu}_e$ similarly escapes the core. The nucleus can now endlessly repeat this sequence whereby escaping neutrinos drain the core of energy. For a nucleus containing N neutrons and Z protons, written as (N, Z), the two-step process is represented by the following reactions:

$$(N, Z) + e^- \;\rightarrow\; (N + 1, Z - 1) + \nu_e \qquad \text{(nuclear electron capture) , and}$$
$$(N + 1, Z - 1) \;\rightarrow\; (N, Z) + e^- + \bar{\nu}_e \qquad \text{(nuclear beta decay) .}$$

During a conference in Urca, Brazil, physicists George Gamow and Mario Schoenberg noted that the local casino appeared to drain money from gamblers much in the way these reactions drained energy from a star. The two physicists promptly dubbed this set of reactions the Urca process.

Although a heavy star will burn hydrogen and helium for many millions of years, it will burn carbon for about 100,000 years and oxygen for only 10,000 years. Incredibly, silicon burning lasts but one day.

Silicon fuses to become iron, but once iron is created, the process of liberating energy through thermonuclear fusion comes to an abrupt end. Iron is the most stable of all nuclei. Any fusion or fission reactions in which iron participates absorb rather than release energy. Thus, formation of the iron core marks the beginning of the end for massive stars. As energy continues to leak out, the core pressure drops, and the core rapidly loses its internal support. The core physically implodes as gravity causes the planet-sized center to collapse under its own weight. As discussed in the next section, the debacle is over in less than one second or, literally, within the blink of an eye.

The Core Collapse

Just prior to its collapse, the silicon and iron core has a radius of about 4,000 kilometers and a mass of about 1.4$M_\odot$. Once silicon burning ends, the core begins to contract. But two events will quickly turn the contraction into a nearly free-fall collapse.

First, compression causes the temperature in the central region of the core to rise above 5 billion kelvins, or 0.5 million electron volts (MeV) of energy per particle. At that temperature, scores of photons generated within the central core are energetic enough to dissociate iron into helium nuclei and neutrons. It was the fusion of those same light nuclei that allowed the star to continually emit energy during the eons of its life. The energy of gravity now undoes that work, as nuclear absorption of a photon breaks the iron apart and sucks thermal energy from the central core.

Second, because the core density has also been steadily rising, the core electrons condense into a special quantum state known as the degenerate electron gas. (See the box "An Exotic State" on the facing page.) Above approximately 10^{10} grams per cubic centimeter (g/cm^3), some of the electrons (e^-) in that unusual state acquire the 2.25 MeV of energy needed to transform free, unbound protons (p) into neutrons (n):

$$p + e^- \rightarrow n + \nu_e \; .$$
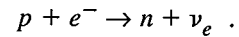
This weak interaction process, called *electron capture*, produces an electron neutrino ν_e that escapes and removes energy. (Note the similarity between this reaction and the nuclear electron-capture reaction discussed in the box "The Urca Process" on this page.)

The probability for electron capture to occur depends on the square of the electron energy. In turn, electrons become more energetic as the core contracts. Higher densities also make encounters between free protons and electrons more and more likely. Thus, as the collapse continues, the rate of the reaction begins to increase dramatically. Free protons and an equal number of electrons disappear as neutrons are produced. The core becomes partially "neutronized." Energy-sapping neutrinos are produced in copious numbers. Despite a rising density and temperature, the rate of cooling inside the central core increases in a runaway fashion as the core implodes.

Along with an increasing cooling rate, the core experiences an ever increasing gravitational force. The strength of gravity varies as $1/r^2$, where r is the radius. As the core shrinks, the gravitational force crushing the core simply gets stronger. The core collapses faster and faster and faster!

Indeed, the collapse would continue indefinitely and create a black hole, if another special quantum state—the degenerate nucleon gas—did not form. A nucleon is either a proton or a neutron. At high densities, the nucleons in the degenerate gas exert substantial pressure and resist being squeezed together. Furthermore, once the central core density surpasses about 10^{14} g/cm^3, nucleons are squeezed so close to one another that very short range, repulsive, nuclear forces come into play. They provide pressure support in addition to that coming from the nucleon degeneracy. The total internal pressure inside the core starts to increase dramatically, and once the

An Exotic State

The incredibly high densities achieved in the stellar core create an exotic form of matter called a degenerate Fermi gas, in which the laws of quantum mechanics hold sway on a macroscopic scale. This gas forms from a set of identical fermions—particles with half-integer intrinsic spin values, such as electrons, protons, neutrons, or neutrinos. The particles in the gas obey the famous Pauli exclusion principle, which states that identical fermions must at all times occupy their own, unique quantum state.* Because states are defined by discrete momentum values, the exclusion principle demands that every particle have a unique momentum and hence a distinct energy.

In an ordinary, classical gas, particles occupy energy states that are distributed about the mean thermal energy of the gas. Typically, most of the low-energy states are unoccupied. But when fermions are forced into such close contact that the exclusion principle applies, a degenerate Fermi gas can form. In that case, particles occupy the lowest possible energy levels and fill states sequentially. This means that the particles are essentially "locked" in their states. They cannot move to lower levels because all lower states are filled. Thus, individual particles cannot lower their energy. Whereas an ordinary gas dissipates energy when particles scatter or radiate photons, the degenerate gas only loses energy by way of particle loss.

A degenerate gas therefore contains a "degeneracy" energy that is largely independent of the thermal energy. But the degeneracy energy grows rapidly with density because each new particle is forced to occupy an unfilled state, and those states always have higher energies. In the superdense core, the degeneracy energy of the gas is enormous—much higher than the thermal energy. Because these arguments also apply to momentum states, and the momentum of particles in a gas relates to the pressure, a degenerate gas exerts a substantial degeneracy pressure that similarly grows with density in a temperature-independent way.

In the core, the electrons begin to form a degenerate gas during the late burning stages. This process boosts the electron energies well above thermal energies and gives them the 0.25 MeV that is needed to drive the Urca process. After the core begins to collapse and the density increases, degeneracy pushes electron energies above the 2.25 MeV threshold of the electron capture process. Finally, it is the growing pressure from the degenerate nucleon gas, formed above 10^{14} g/cm^3, that ultimately halts the collapse of the core.

---

*Because there are two spin states (up or down), two particles can occupy each state. The basic discussion does not change.

central density surpasses about 3×10^{14} g/cm^3, the pressure becomes so great that the very center of the collapsing iron core—a 10-kilometer-radius "inner core" of unimaginably high temperature and density—effectively becomes incompressible. Its collapse abruptly halts.

Something remarkable then occurs. Much like an overcompressed rubber ball that is suddenly allowed to return to equilibrium, the inner core violently reexpands. A layer of dense matter surges outward at roughly 10,000 kilometers per second, and a strong shock front begins to plow through material that is still falling inward at roughly 60,000 kilometers per second. This dynamic event is often referred to as the *core bounce.*

As explained in the section "Making

Stars Explode" on page 171, the shock front quickly loses energy, and the rapid expansion stops. But gravity cannot cause the dense layer of material behind the shock front to recollapse. The pressure of the degenerate electron gas is very strong and can support that matter out to large radii. (Neutrinos also help provide pressure support. That atypical neutrino behavior will also be discussed in detail later.) Thus, a relatively static layer of hot, dense matter—less dense than the inner core but much more dense than the material that continues to fall—forms as a result of the core bounce. This layer grows larger as new infall adds to it. The shock front, which demarcates this high-density layer from the low-density infall, begins to slowly move out. These events are summarized in Figure 3.
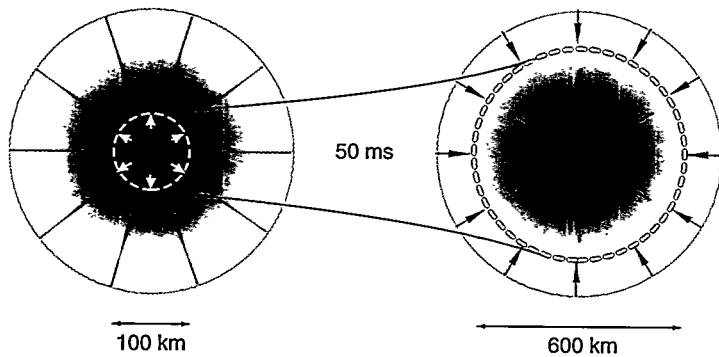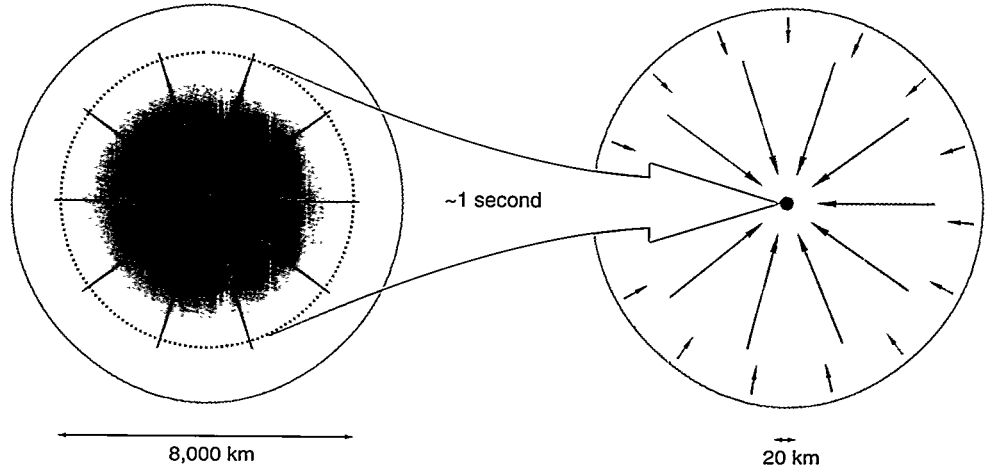
The central portion of that dense layer, out to a radius of approximately 40 kilometers and including the inner core, plays a critical role in the formation of supernovae. That region is called the proto-neutron star (refer again to Figure 3). The proto-neutron star is about $1.2M_\odot$ of electrons, neutrinos, and nuclear matter. After about 10 seconds, it will cool and condense to become the much denser neutron star that is the endpoint of a massive star's evolution. But the supernova explosion is initiated well before the neutron star forms. Thus, the focus of supernova physics is on the development of the proto-neutron star and on the processes that occur in the quasi-static layer of matter that develops behind the slowly expanding shock front.

Before we move on, let's take a minute to savor one of nature's most sublime moments. Prior to its collapse, the stellar core is more than twice as large as the moon (although it is about 50 million times heavier!). *In less than one second,* nearly one-third of that mass is compacted into a sphere that would easily fit inside the city of St. Louis, Missouri. The collapse happens so quickly that the remainder of the original iron core, which still extends out to a radius of a few thousand

Figure 3. Core Collapse, Bounce, and the Postbounce Structure of the New Core

Once silicon burning ends, the star's core rapidly cools and loses pressure. As shown in (a), the entire 8,000-kilometer core undergoes a nearly free-fall collapse and, in less than 1 second, about one-third of the mass is crushed into a sphere approximately 20 kilometers in diameter. The new core, shown in (b) and (c), is vastly denser than the original. In these illustrations, arrows represent moving matter, and the length of the arrow is generally indicative of velocity.

(a) Core collapse. Core material (enclosed by red dotted line) races inward (long arrows), slowing down only when it runs into the dense matter in the center. The collapse happens so quickly that most of the material outside the core is unaware that the collapse has even occurred. The original core had an average density of about 10^7 g/cm^3, but the average density of the new core (red dot) exceeds 10^{14} g/cm^3, which is more than 10 million times greater than before.

~1 second

8,000 km   20 km

50 ms

100 km   600 km

10^{14}

Inner core

Proto-neutron star

10^{12}

Density (g/cm^3)

Quasi-static layer

10^{10}

Shock front

10^8

10^0 10^1 10^2 10^3

Radius (km)

(b) Core bounce. Once the central density of the new core reaches 3×10^{14} g/cm^3, the "inner" core effectively becomes incompressible and no longer collapses. Immediately, the inner core expands and surges outward (white arrows). It quickly swells to a diameter of about 80 kilometers (the proto-neutron star, shown in gray) as a high-velocity shock front (white, dashed line), pushes its way against the infall. Within a few milliseconds, however, the shock loses energy, stalls, and stops its rapid outward movement.
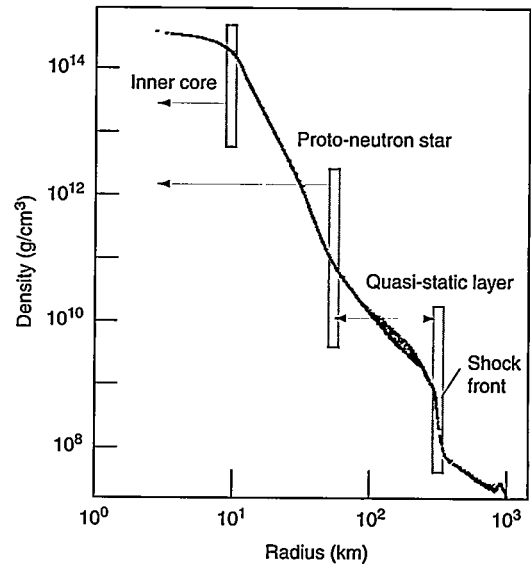
(c) Postbounce structure and density profile of the core. As infall passes through the shock front, the material slows down, and its density increases dramatically (small, black arrows). A dense layer, called the quasi-static layer, begins to accumulate outside the proto-neutron star. The shock front slowly moves outward. The graph of density versus radius (both plotted on log scales) reveals the extent of the quasi-static layer, the proto-neutron star, and the inner core (the very center of the proto-neutron star) 50 milliseconds after the bounce. These regions are typically defined in terms of density. At roughly 3×10^{14} g/cm^3, the inner core extends to a radius of about 10 kilometers and encloses about $0.5 M_\odot$. The proto-neutron star decreases in density to about 10^{11} g/cm^3, encloses about $1.2 M_\odot$, and extends to about 40 kilometers. The shock front (evident as a sharp change in density) has by this time advanced to an approximate radius of 300 kilometers.

kilometers, is left trying to catch up. This outer region—an iron shell of comparatively low density and internal pressure—offers little resistance to the inward pull of gravity. It is racing toward the proto-neutron star at the incredible velocity of 60,000 kilometers per second.

Finally, most of the star is completely unaware of what is going on. Information about the collapse of the core can only travel as fast as the speed of sound in gas. In the fraction of a second during which the collapse occurs, information can only reach out to a few thousand kilometers. The stellar envelope may extend tens of millions to hundreds of millions of kilometers into space. Thus, much like a cartoon character suspended in midair, most of the star has yet to learn that the rug has been pulled out from under it.

The physics of events leading up to and immediately following core bounce has been fairly well understood for the last twenty-five years, although the models underwent many revisions and modifications as more processes were considered and the role of neutrinos became clearer. But a consensus regarding the postbounce physics many milliseconds after the bounce, the crux of Type II supernovae dynamics, has been much slower to emerge. In a nutshell, the problem is how to turn an implosion into an explosion.

Making Stars Explode

The first modern model of supernovae was presented in 1960 by Stirling Colgate and Montgomery Johnson. It postulates that the outward-moving shock wave produced by the core bounce is sufficiently energetic to continue moving through the outer core like a sonic boom. The shock eventually expels the stellar envelope in a large explosion. This model later became known as the "prompt" mechanism because it argues that the explosion occurs immediately after the bounce.

But in order to continue propagating, the shock needs to beat back the infall of the rest of the star. The postshock temperature is so high, however, that many of the cooling processes that initially led to the collapse of the core, namely, iron photodisintegration and intense neutrino emission, apply equally well to the shock front. As alluded to in the previous section, the shock stalls for all but the most extreme assumptions about the precollapse structure of the star. The bounce shock is thus unable to deliver an explosion and halt the infall of the stellar envelope. In the prompt model, the shock front retreats, and the massive star collapses to a black hole.

Around the same time that prompt models were being developed, scientists realized that there was far more energy available to power supernova explosions than what was typically measured as the kinetic energy of the debris. Based on observations, the explosive energies of supernovae typically tally to about 10^{51} ergs, or 1 "foe." Hans Bethe coined the acronym for (ten to the) fifty-one ergs. But the entire mass of the precollapsed core eventually ends up in a neutron star whose radius is only 10 kilometers, or one sixty-millionth of the original core's radius! The work done by gravity in compressing the core represents a total energy on the order of 300 foe, or nearly 300 times more energy than what is typically observed to be released by the explosion.

It was recognized that most of the energy is carried off by neutrinos that are created as the core becomes neutronized through electron capture.

Today, it seems natural to expect that a small fraction of that energetic neutrino flux powers supernovae. However, in the early 1960s, the idea that neutrinos might do anything dynamical, let alone power an explosion, seemed preposterous.

It was in this context that in 1965 Stirling Colgate and Richard White put forth the first model invoking heating by neutrinos as the mechanism responsible for supernovae. They used a hydrodynamic code to quantitatively analyze their theory. Theirs was the first attempt to simulate the hydrodynamics of a supernova. It was probably the first hydrodynamic simulation ever done in astrophysics.

According to Colgate and White, a supernova is initiated when an iron core collapses directly to a neutron star. As falling matter collides with this very small, incompressible object, a shock front develops that is hot enough to emit neutrinos. Falling material absorbs the neutrinos, heats up, and expands. A mighty explosion ensues.

But the Colgate and White model was eventually shown not to work.[4] It failed, in part, because of a missing piece of neutrino physics that was neither experimentally confirmed nor appreciated until the mid-70s. That missing piece was the neutrino neutral-current scattering.

Neutral-current scattering was a new type of neutrino interaction, and at the high densities found within the core, it resulted in the efficient scattering of neutrinos from nuclei and unbound nucleons. The neutrinos would no

---

[4]The Colgate and White model is frequently believed to have failed because of an improper postbounce neutrino emission and absorption algorithm. In fact, the model fails because it neglects the effects of neutrino cooling during core collapse. In the model, a supernova develops when a high-temperature iron core collapses. The high temperature leads to a very rapid rate of electron capture, and the core becomes neutronized very quickly and at relatively low densities. A neutron star forms directly from the collapse. The collision energy of the infall onto the neutron star is high enough to generate a high-temperature accretion shock front, and high-energy neutrinos emitted from that front are readily absorbed in the falling matter. Once neutrino cooling was added to the model, the core temperature and hence the nucleon boiloff rate were reduced relative to what Colgate and White had originally considered. Neutronization proceeded much more slowly because there were fewer free protons. In addition, neutral-current interactions were not known at the time. They had the effect of enhancing the neutrino trapping rate, which further retarded core neutronization. Neutrino trapping also led to a large degenerate lepton pressure that supported matter at a radius some 3 to 5 times greater than the radius of a neutron star. Thus, it was eventually shown that the accretion shock front formed at larger radii, and matter that fell onto this shock front was not nearly energetic enough to produce the high-energy neutrinos needed to drive a supernova.

what is called the lepton degeneracy pressure) that support the shock front and prevent gravitational collapse.

However, even with neutrino trapping incorporated into the models, efforts to obtain explosions were frequently thwarted. Stellar fizzles were often the result of a detailed calculation. But a major shift in supernova models occurred in 1982, when James Wilson began running computer simulations that tracked events over very long periods of time. Partly because of computer limitations, researchers had tended to model only the core collapse and the events that occurred a few tens of milliseconds after the bounce. Wilson's simulations ran from the start of core collapse to about half a second after the bounce. In his simulations, apparent fizzles evolved into successful blowouts by what later was called the "delayed" (as opposed to prompt) mechanism.

In both the prompt and delayed models, the bounce shock moves out a few hundred kilometers beyond the proto-neutron star and stalls. A stagnant shock front would normally be a sign that all outward expansion has stopped, in which case no prompt explosion occurs and the star inevitably recollapses to a black hole.

But the bounce shock does play a crucial role in setting the stage for the success of the delayed mechanism. After the bounce shock stalls, the degenerate lepton pressure prevents material from recollapsing directly onto the proto-neutron star. By tracking the physics for long periods of time, the simulation showed that the shock front is able to withstand the initially large ram pressure of the infall and is still present when that pressure begins to subside. As a result, the quasi-static layer between the stalled shock and the surface of the proto-neutron star persists longer than the neutrino-diffusion time scale. Some of the energetic neutrinos slowly leaking out of the proto-neutron star can be absorbed in the dense material behind the shock front. Material is constantly heated

Neutrino Trapping

The neutrino is the particle that embodies the weak interactions. Up until 1973, neutrinos had been observed to participate only in charge-changing weak interactions, such as electron capture or the reactions making up the two-step Urca process. Two interacting particles exchange a W^+ or W^- boson, and so exchange one unit of electric charge. Charge-changing reactions occur so infrequently that, even at the high densities reached during core collapse, the neutrinos were thought to simply free-stream out of the core.

But in 1973 the neutral-current interaction, long predicted by theorists to be a necessary consequence of electroweak unification, was experimentally verified. This was a new type of weak interaction in which particles exchange a Z^0 boson. Thus, there is no change in the charge states of the participants. Instead, a neutrino could merely scatter from nucleons or electrons. In 1975, Tubbs and Schramm found neutral-current scattering to be favored under the conditions prevailing during core collapse. The neutrino could simultaneously scatter from all the nucleons in a heavy nucleus in a coherent process that boosted the scattering cross section by more than 1 order of magnitude over charged-current processes. At densities above 10^{11} g/cm^3, neutrinos began to scatter from nuclei so often that they became trapped within the core.

One profound consequence of the trapping is that the neutrino density increases enough to reverse the direction of the electron capture reaction:

$$p + e^- \leftrightarrow n + \nu_e .$$

Neutrons are transformed back into protons, thus allowing a proton/neutron equilibrium to be established. Neutron star formation is inhibited, and the proto-neutron star forms instead. A second consequence of the trapping is that the neutrino stays in the core long enough to form a degenerate gas. Together with electrons, the two light particles form a *degenerate lepton gas*. It is the lepton gas that stores most of the energy liberated by the gravitational collapse of the core, and it is also the lepton degeneracy pressure that expands the proto-neutron star and supports the bounce shock front long after core bounce has occurred.[*] Neutrinos of all flavors will scatter via neutral-current interactions, so that ν_μ and ν_τ neutrinos, produced as the core collapses, are also trapped.

[*] Note that the degenerate lepton pressure is unable to halt the initial collapse of the core. The response of the relativistic lepton gas to further compression is "mushy," and the pressure does not increase very fast when the gas is compressed. The strength of gravity, however, increases nonlinearly with decreasing radius, and the lepton degeneracy pressure alone is insufficient to overcome the increasing pull of gravity as the collapse proceeds.

longer escape blithely from the super-dense proto-neutron star but would instead become "trapped" and take several seconds to escape (see the box "Neutrino Trapping" on this page). Indeed, neutrino trapping can be used to "define" the proto-neutron star, in that inside the proto-neutron star, neutrinos are trapped. Outside the proto-neutron star, neutrinos no longer scatter strongly but free-stream through the star.

In many ways, neutrino trapping was remarkable. A neutrino is a particle that ordinarily passes through *half a light-year* of lead without scattering! But for a few seconds in the center of a dying star, neutrinos behave like any other particle. They scatter, are constantly absorbed and reemitted, and significantly, exert degeneracy pressure. It is the neutrino and electron degeneracy pressures (the dominant components of

from below, and it can expand and continue to drive the shock front outward (see Figure 4).

Although delayed models could produce explosions, less satisfying was the fact that, all through the 70s and 80s, supernova simulations seemed to be highly sensitive to the smallest details of how the physics was implemented. Whereas one group might obtain explosions, another would get fizzles simply because the approximations used in the modeling were different. This was worrisome not only because it put any calculations at the mercy of a new wrinkle in the theory, but also because real supernovae do not seem to have such problems. Explosions of fairly uniform energies, always of the order of 1 foe, appear to be produced quite readily. Supernova theory seemed trapped by an endless cycle of successful and failed explosions. Sheer desperation led astrophysicists to consider numerous alternative theories involving core rotation, nuclear burning, magnetic fields, and other processes. However, none of these worked well. What astrophysicists really needed was some sort of lucky break, and on February 24, 1987, they got it.

Region cooled by neutrinos
40- to 50-km radius; neutrinos are produced by the following reactions:

$$p + e^- \rightarrow n + \nu_e$$
$$n + e^+ \rightarrow p + \bar{\nu}_e$$
$$e^+ + e^- \rightarrow \nu + \bar{\nu}$$

Region heated by neutrinos
50- to 100-km radius; neutrinos are absorbed by the following reactions:

$$n + \nu_e \rightarrow p + e^-$$
$$p + \bar{\nu}_e \rightarrow n + e^+$$
$$\nu + e^{+/-} \rightarrow \nu + e^{+/-}$$

Rapidly falling material

Quasi-static layer

Stalled shock front

Hot neutrinos

Proto-neutron star approximately 40-km radius

Figure 4. Behind the Front: Heating Matter with Neutrinos
As a result of core bounce, a shock front moves beyond the ultradense surface of the proto-neutron star. The shock loses energy as it propagates and stalls. It is prevented from recollapsing by the pressure support of the degenerate lepton gas, and so it remains at a relatively stable radius, creating a quasi-static layer of dense matter. Some of the energetic neutrinos leaking from the proto-neutron star deposit their energy in the quasi-static layer. The matter expands and becomes buoyant. The neutrinos, therefore, transfer energy out of the extremely high temperature core and into a large mass of lower-temperature material.

Supernova 1987A

Supernova 1987A (SN1987A), the first supernova seen in 1987, owes its major impact on supernova theory to one reason: it occurred relatively nearby. It flared up a modest 170,000 light-years away in the Large Magellanic Cloud, which is a satellite galaxy of our own Milky Way galaxy. (See the box "Supernova 1987A" on the facing page.) For the first time, it was possible to look back in photographic archives at the location of the explosion and find the parent star of the supernova, which was a $20M_\odot$ blue supergiant. Because of the star's proximity and, hence, the brightness of the supernova, observations of unprecedented accuracy became possible. (It should not be overlooked that
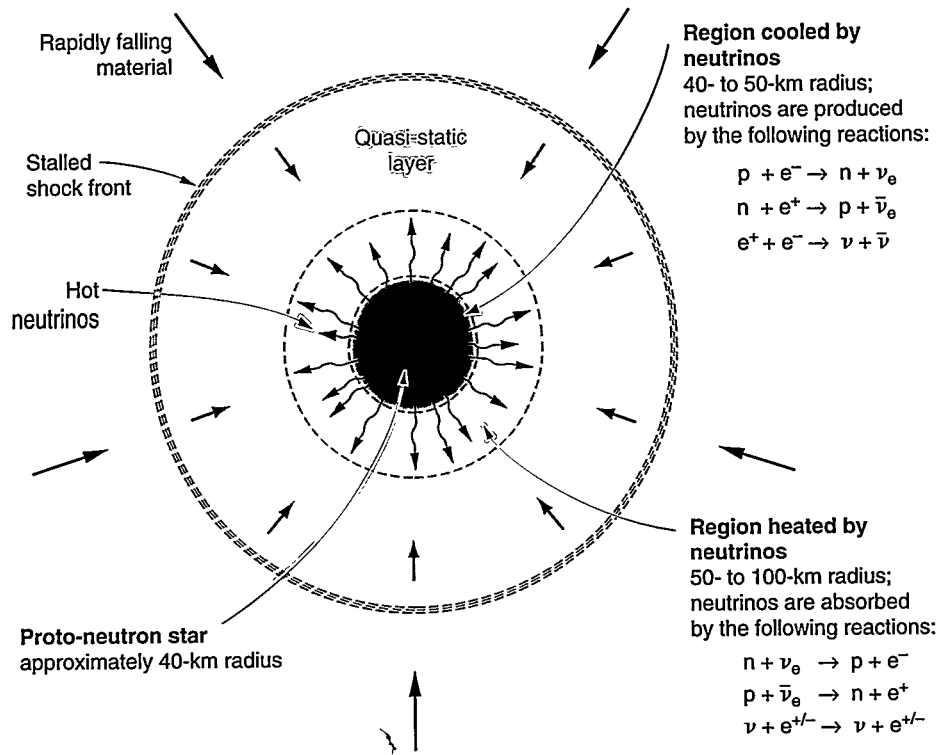
SN1987A occurred during the current "golden age" of astronomy, when numerous observatories worldwide have sophisticated equipment in place.)

Most important, however, SN1987A is the only supernova from which neutrinos were observed. Two underground detectors sensitive to electron antineutrinos, Kamiokande II in Japan and IMB in Ohio, detected bursts of twelve and eight antineutrinos, respectively, over a 10-second interval. The small number of events did not allow for detailed quantitative modeling of SN1987A, but it did provide qualitative estimates of what had happened.

The detected signal strongly supports the picture of a hot proto-neutron star forming and cooling by neutrino emission and is entirely consistent

with our current theories of core collapse. The energies of individual neutrinos correspond to the expected initial temperature of a proto-neutron star, while the duration of the bursts is in line with the 10-second cooling time for such an object. The energy spectrum of the neutrinos permitted an estimate of the total energy radiated during the supernova, which is consistent with the creation of a $1.4M_\odot$ neutron star whose radius measures 15 kilometers.

At the same time, analysis of the emission spectra of SN1987A unequivocally showed that the ejected envelope was stirred up considerably during the explosion. Especially puzzling was the presence of iron in the outer hydrogen and helium layers of the ejecta, indicating that a substantial amount of mixing
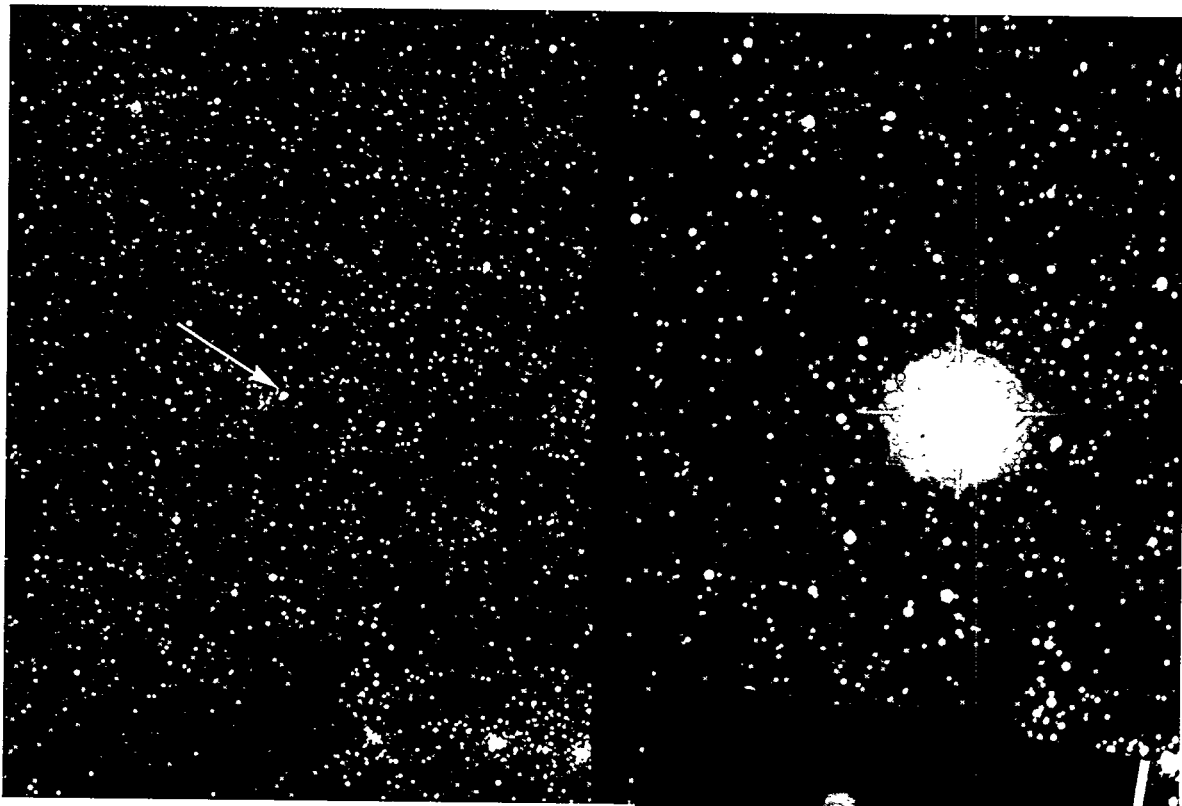
Supernova 1987A
a brief photo history
On February 24, 1987, the astronomy community was
startled and delighted by the appearance of a dazzling
supernova in the Large Magellanic Cloud, which is a
companion galaxy to our own Milky Way galaxy and
is visible from the southern hemisphere. The super-
bright "new star" could be easily seen by the naked
eye. In the pair of photos shown at the top of the
page, the arrow in the photo on the left points to a
$20M_\odot$ blue supergiant. The photo was taken in
February 1984. The photo to its right was taken on March 8 and 9,
1987, with the 3.9-meter Anglo-Australian telescope at the Anglo-Australian Observatory (in
New South Wales, Australia). The star has become a supernova.
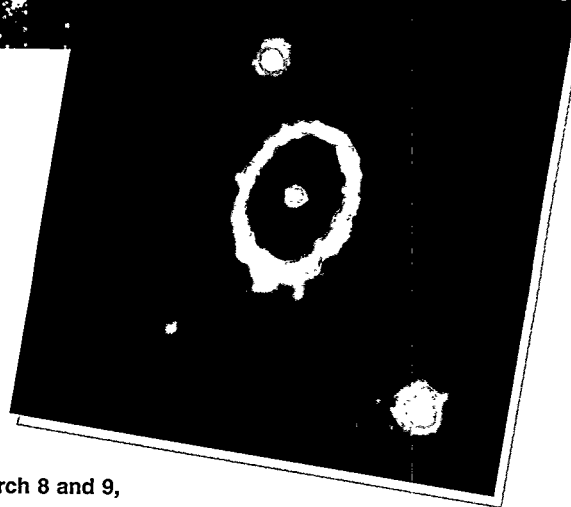
Seven years later, in the spring of 1994, the Hubble Space Telescope trained on-site its wide-field planetary
camera 2 to record the three-ring structure pictured above. The rings are most likely in three parallel but
separate planes that are inclined to our point of view, making the rings appear to intersect. The small,
bright central ring surrounds the supernova site, and the two larger rings are presumably lying in front of
and behind the site.

(Left and middle photos—© Anglo-Australian Observatory; Right photo Dr. C. Burrows, ESA/STScI and NASA, press release #STScI-PR94-22, created
with support to Space Telescope Science Institute, operated by the Association of Universities for Research in Astronomy, Inc., from NASA contract
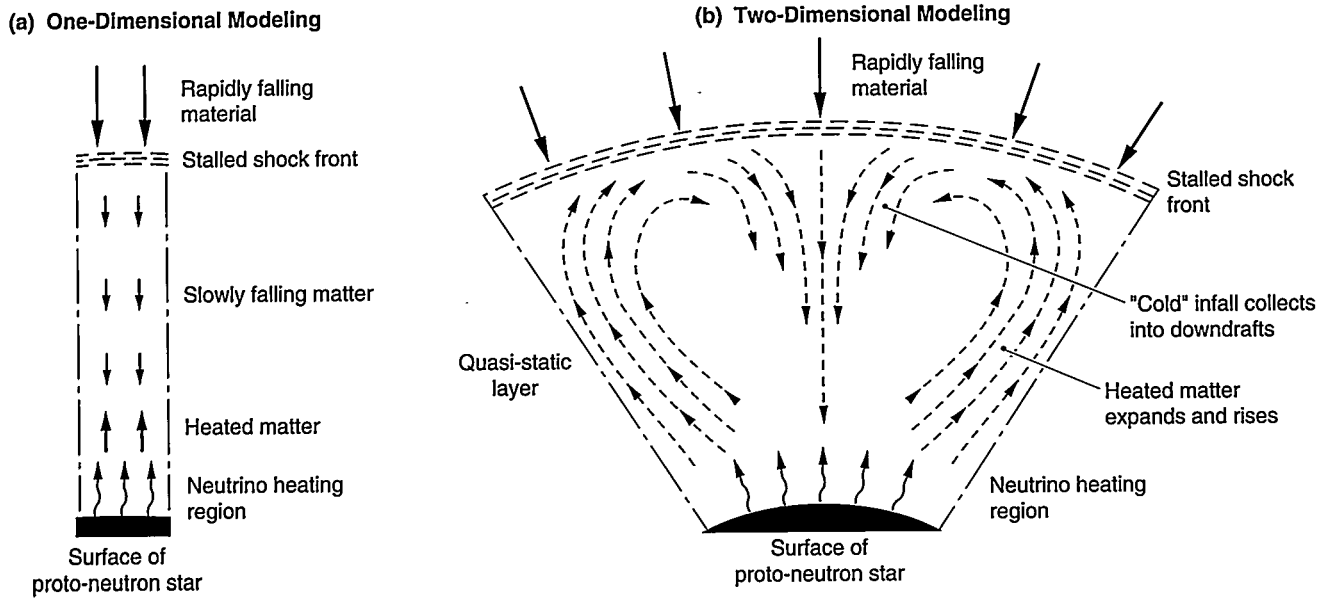NAS 5-26555. Reproduced with permission from AURA/STScI.)

(a) One-Dimensional Modeling

Rapidly falling material

Stalled shock front

Slowly falling matter

Heated matter

Neutrino heating region

Surface of proto-neutron star

(b) Two-Dimensional Modeling

Rapidly falling material

Stalled shock front

"Cold" infall collects into downdrafts

Heated matter expands and rises

Quasi-static layer

Neutrino heating region

Surface of proto-neutron star

Figure 5. A Convective Engine

(a) For simplicity, supernovae were often modeled in one dimension. A star was assumed to be spherically symmetric, its radius being the only spatial parameter that mattered. Doing simulations was therefore equivalent to doing physics in a long tube, even though the transfer of heat from one end of a pipe to the other is not very effective. (b) With the advent of multidimensional models, convection could occur. Hot, buoyant material could rise in one part of the star, to be replaced by cooler material falling from some other region. An in-out circuit is established that allows for the efficient and continuous transfer of heat out of the core and into the quasi-static layer. Energy from the gravitational collapse is thus converted into mechanical work as heat is being transferred between hot and cold reservoirs. In this sense, supernovae can be thought of as being powered by a simple convective engine.

had taken place over very large distances (tens of millions of kilometers). Some of this mixing was explained by instabilities that occurred while the shock wave was running from the core to the distant surface of the star, well after the explosion had been launched. Nevertheless, these observations promoted an awareness that violent instabilities might be involved in the explosion mechanism.

This idea was not entirely new. In 1979, Richard Epstein of Los Alamos Scientific Laboratory had already proposed that instabilities at the edge of the proto-neutron star might be important. Hans Bethe later pointed out that an explosion due to neutrino heating, as in the delayed mechanism, would necessarily lead to convection because matter is "heated from below."

However, computer limitations and the complexity of supernova physics led most astrophysicists to simplify simulations by assuming spherical

symmetry. The problem was therefore reduced to one spatial dimension—the radius. As a result, instabilities were thought to mix matter at microscopic scales; they were not thought to lead to large-scale bulk flows.

Over the years that followed the advent of SN1987A, we became convinced that to explain the observed churning of elements, one had to look into the explosion mechanism. We felt that "standard" one-dimensional modeling was likely to miss some important qualitative aspects that followed core collapse. As a result, in 1991 we started research with the goal of simulating the explosion mechanism in multidimensions. Of great help were newly available, inexpensive, powerful desktop computers on which two-dimensional and sometime three-dimensional simulations could be run.

Even from primitive, initial calculations, we noticed intense convective instabilities (akin to boiling) arising

from the simultaneous existence of cold inflows and heated outflows. The convection was driven by matter made hot and buoyant by neutrino heating. Such instabilities were impossible to model by one-dimensional simulations that average quantities at a given radius (Figure 5).

Figure 6 on the next page is a snapshot of the core region 50 milliseconds after core bounce. As in other models, our postbounce shock wave is stalled and is now at a radius of about 300 kilometers. As falling matter passes through the shock front, its density increases, and its velocity decreases. The matter meets with larger and larger neutrino fluxes, is heated, and expands into large bubbles that rise through the quasi-static layer like hot-air balloons. The bubbles push against the shock. As time passes (Figure 7), more and more bubbles collect and push until the shock is finally driven outward. *The star becomes a supernova!*

Figure 6. Computer Simulation of Neutrino-Driven Convection
This graphic shows a slice of the core region 50 milliseconds after the bounce. Parcels of matter are shown as colored arrows; the length and direction of an arrow indicates velocity, and color indicates entropy (*S*). Regions of higher entropy correspond to regions that have been heated. The shock front (where yellow arrows meet green arrows) lies at about 300 kilometers. Low-entropy, high-velocity material (blue arrows) rains down on the shock, and its entropy increases as it moves through the front. The material becomes part of the quasi-static layer. Closer to the core, energetic neutrinos streaming out of the proto-neutron star (blue-green region extending to about 40 kilometers) are absorbed in the quasi-static layer, which becomes heated. High-entropy bubbles (red) are already rising. They will transfer energy to the shock front, reenergizing it and allowing it to move farther out. Low-entropy downdrafts have formed (yellow filaments) that funnel cooler material toward the proto-neutron star, thus closing the convective loop.
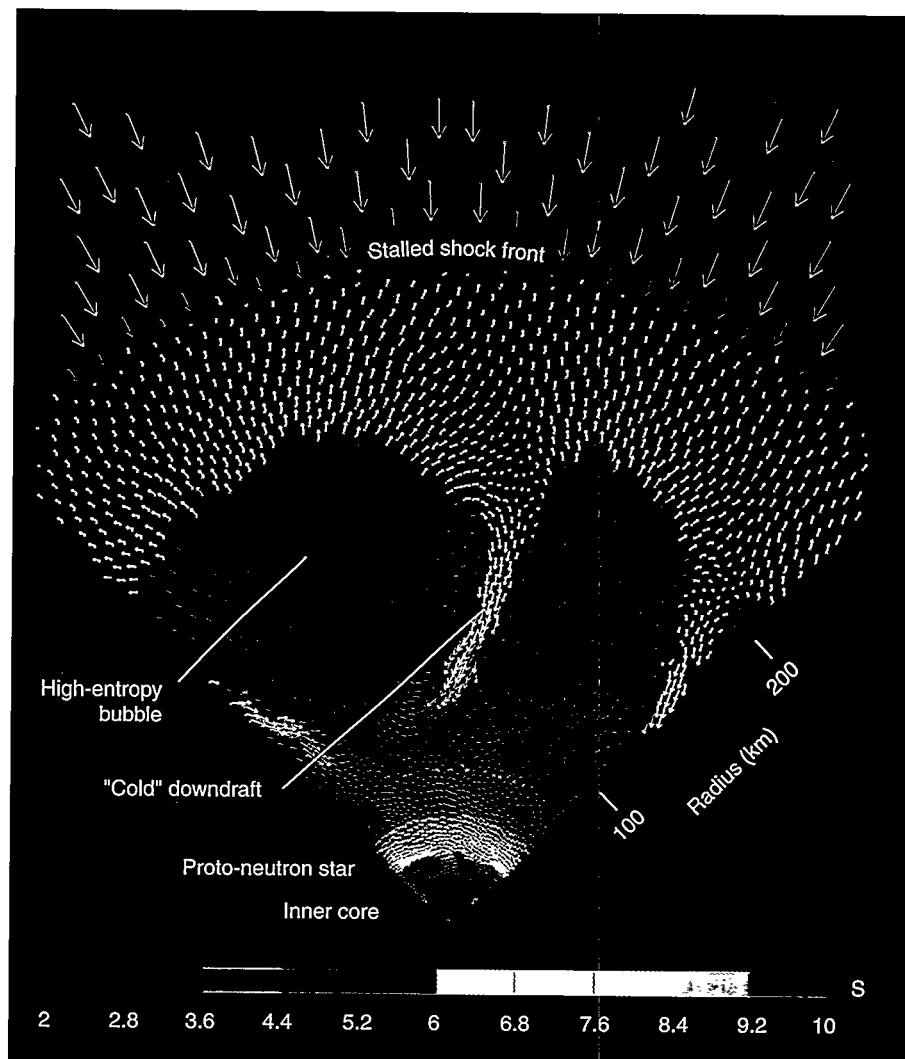
Supernovae and Convection

Obtaining a supernova explosion is somewhat akin to blowing up an ordinary pressure cooker. The lid of the cooker is the ram pressure of the falling matter; the stove is a hot proto-neutron star. Blowing up the cooker requires a buildup of pressure against the lid, which in turn depends on a good transport of heat between the top and bottom of the cooker. It is convection that allows heat to be carried to the lid. The pressure builds up until the lid finally pops.

In more physical terms, our simulations led us to elaborate on a new paradigm in which the supernova is viewed as a convective engine.

The proto-neutron star is viewed as a heat source radiating neutrinos, and the envelope of the star is a cold reservoir. The circulation of matter and the exchange of heat allow mechanical work to be extracted from the energy liberated by the gravitational collapse (see Figure 8). This paradigm explains the failure or marginality of simulations in one dimension; heat transport with one pipe can hardly be effective. But in two dimensions, an in-out circuit can be established.

The transport of energy via convection has the additional feature that the explosive energies are self-limiting. Once an explosion occurs, matter is ejected and dispersed into a nebulous cloud of gas. There is no

more matter left to heat, and the energy input stops. Thus, the model arrives at a natural explanation of the general constancy of explosion energies for different supernovae.

Furthermore, our simulations were very encouraging because successful explosions were obtained in a way that seemed fairly insensitive to the details of the numerical implementation of the physics. Subsequent, increasingly realistic simulations (that is, simulations that tracked more physical processes) by us and others confirmed the key role of neutrino-driven convection in the genesis of the explosion.

Despite the success of our model, current multidimensional simulations still have significant problems.
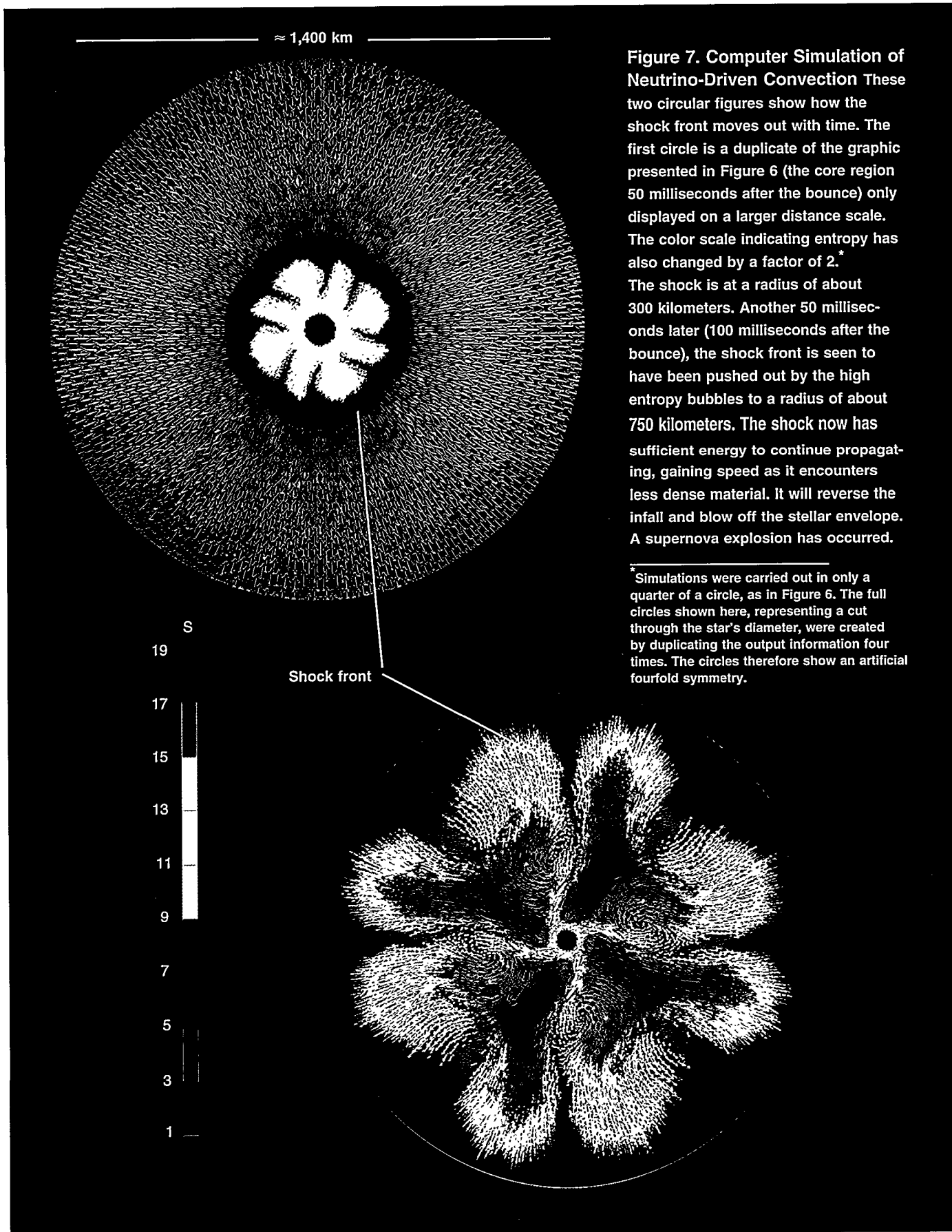
≈ 1,400 km

Figure 7. Computer Simulation of Neutrino-Driven Convection These two circular figures show how the shock front moves out with time. The first circle is a duplicate of the graphic presented in Figure 6 (the core region 50 milliseconds after the bounce) only displayed on a larger distance scale. The color scale indicating entropy has also changed by a factor of 2.*
The shock is at a radius of about 300 kilometers. Another 50 milliseconds later (100 milliseconds after the bounce), the shock front is seen to have been pushed out by the high entropy bubbles to a radius of about 750 kilometers. The shock now has sufficient energy to continue propagating, gaining speed as it encounters less dense material. It will reverse the infall and blow off the stellar envelope. A supernova explosion has occurred.

*Simulations were carried out in only a quarter of a circle, as in Figure 6. The full circles shown here, representing a cut through the star's diameter, were created by duplicating the output information four times. The circles therefore show an artificial fourfold symmetry.

S

19

17

15

13

11

9

7

5

3

1

Shock front

**Figure 8. A Thermodynamic
Heat Cycle**
Shown here is the distribution of core
particles involved in convection, plotted in
the plane of pressure (*P*) versus specific
volume (*V*) 50 milliseconds after the
bounce. The particles lie along a loop that
essentially corresponds to the path they
follow over time in the *P-V* plane.
Integrating pressure versus volume
around the loop yields the mechanical
energy per unit mass (the work done per
gram of matter) delivered by the convec-
tive engine. A crude estimate yields about
4 foes per solar mass, or about 1.2 foes
for the approximately 0.3$M_\odot$ involved in
the convective cycling.
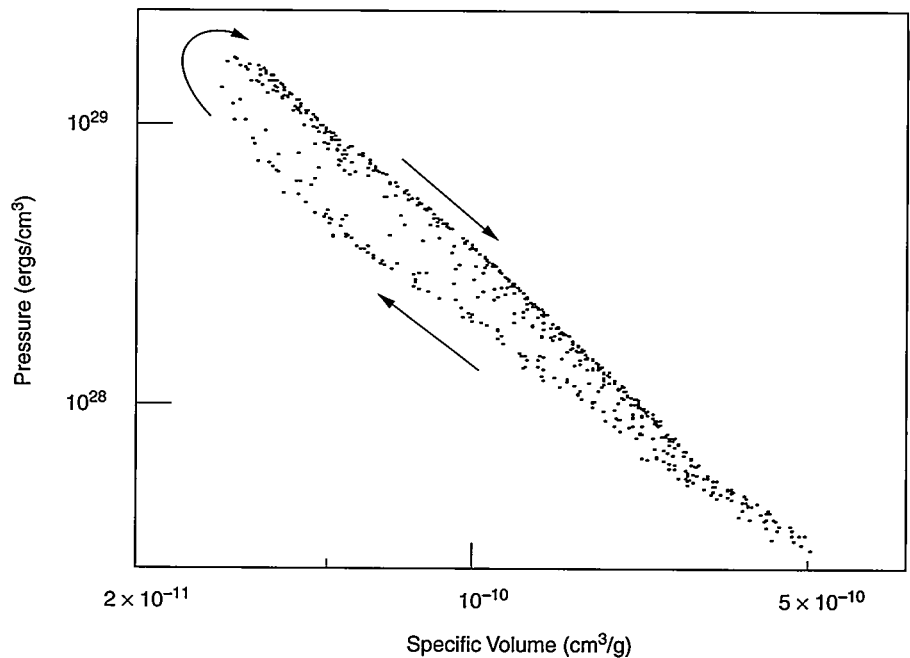
The calculated remnant neutron-star masses are too low when compared with observed masses in neutron star binaries. Also, in comparison with observed solar and terrestrial chemical abundances, the simulation has too much neutron–rich material (such as krypton) being ejected in the explosion.

Some of these problems may be due to the inevitable compromises that had to be made in order to run two- or three-dimensional versus one-dimensional simulations. For instance, the multidimensional scheme to track neutrinos had to be made considerably simpler than the one-dimensional transport algorithms. Similarly, the general relativistic corrections to classical Newtonian gravity are more difficult to implement in multidimensional calculations. These limitations are gradually being overcome, and hopefully, the agreement with observations will improve. Recently, however, researchers using an improved multi-group neutrino diffusion had difficulty obtaining supernovae even after incorporating convection. Could it be that obtaining explosions requires additional physics?

One exciting possibility is that these discrepancies point toward the existence of some new physics beyond the

standard model, such as neutrino oscillations. In the MSW picture, which requires that neutrinos have mass, the enhanced oscillation of one neutrino species into another is triggered by the passage of the neutrino through matter of a certain density. (See the article "MSW" on page 156). Considering that, at the time of collapse, the densities in supernovae range all the way from 10^{14} to 10^{-5} g/cm^3, it is clear that, should neutrinos oscillate, they will most probably do so during supernova explosions.

Of great interest is the density range between 10^{12} and 10^7 g/cm^3. The first density corresponds to the surface of the proto-neutron star, where neutrinos stop diffusing and start free-streaming. The second density corresponds to the outer edge of the neutrino heating region outside the proto-neutron star. Because electron neutrinos are most easily absorbed by nucleons, they are the most efficient at heating. One can envision that tau or muon neutrinos created within the proto-neutron star might oscillate into electron neutrinos between the emission and absorption regions, which would result in more heating than currently predicted. The converse may also be true—electron

neutrinos are lost through oscillations; hence, the heating is reduced. In short, if neutrino oscillations exist, they could have an important impact on the dynamics of the explosion.

The Last Word

One further significance of the neutrino signal from SN1987A is that it placed a new limit on the mass of the electron neutrino. The speed of a massive particle depends on its mass and energy. Because each neutrino let loose by SN1987A traversed the same 170,000 light-years in reaching Earth, one can use the measured spread in neutrino energies and in arrival times to deduce the speed and hence to constrain the neutrino mass. The result is less than 10 electron volts, slightly better than prior experimental limits.

Neutrino and supernova physics are intimately linked. It is therefore not surprising that one of the dearest wishes of astronomers and neutrino physicists alike is for a supernova to occur within our own galaxy. If such an explosion were to take place, it is estimated that the new, large neutrino detectors would register several thousand events. This would provide

us with a detailed picture of the events that accompany the collapse of the core, a picture that is otherwise shielded from our view by the opaque envelope of the star. Moreover, an intense neutrino signal would provide clues and constraints on neutrino oscillations or other physical processes that we may not have imagined yet. It is in part the prospect of such serendipitous discoveries that promises to make the field of supernova and neutrino astrophysics an exciting one for years to come. ■

Further Reading

Benz, W., S. A. Colgate, and M. Herant. 1994. Supernova Explosions and Hydrodynamic Instabilities: From Core Bounce to 90 Days. *Physica D* **77**: 305.

Bethe, H. A. 1990. Supernova Mechanism. *Reviews of Modern Physics* **62**: 801.

Burrows, A., D. Klein, and R. Gandhi. 1992. The Future of Supernova Neutrino Detection. *Physical Review D* **45** (5): 3361.

Colgate, S. A., and M. H. Johnson. 1960. Hydrodynamic Origin of Cosmic Rays. *Physical Review Letters* **5**: 235.

Colgate, S. A., and R. H. White. 1966. Hydrodynamic Behavior of Supernovae Explosions. *Astrophysical Journal* **143**: 626.

Epstein, R. I., Monthly Notices of the Royal Astronomical Society **188**: 305.

Herant, M., W. Benz, W. R. Hix, et al. 1994. Inside the Supernova: A Powerful Convective Engine. *Astrophysical Journal* **435**: 339.

Tubbs, D. L., and D. N. Schramm. 1975. Neutrino Opacities at High Temperatures and Densities. *Astrophysical Journal* **201** (2): 467.

Cantrella, J. M., et al., Eds. 1985. *Numerical Astrophysics*. Boston: Jones & Bartlett.

Woosley, S. E., et al. 1994. The R-Process and Neutrino Heated Supernova Ejecta. *Astrophysical Journal* **433**: 229.

Woosley, S. E. 1996. Nucleosynthesis in Supernovae. *Astronomical Society of the Pacific Conference Series* **99**: 253.

Marc Herant received a B.S. in physics from Caltech. Despite some equivocating about the relative merits of exploring the world and scientific research, Marc entered graduate school in astronomy at Harvard University. After sticking his fingers in a variety of projects, he finally settled on writing a Ph.D. thesis on the hydrodynamics of supernova 1987A (SN1987A) under the benign supervision of his advisor, Willy Benz. He then took advantage of a NASA-sponsored Gamma–Ray Observatory Postdoctoral Fellowship to extend his research on supernovae with Stan Woosley at the University of California, Santa Cruz. This fellowship was followed in 1994 by a Director Postdoctoral Fellowship at Los Alamos, where Marc was able to indulge in his interest in the life sciences by studying the mechanics of cell motility. Marc now lives with his wife Debra in St. Louis, Missouri, where he is a second-year medical student at Washington University.

Stirling A. Colgate earned his B.S. and Ph.D. degrees in physics from Cornell University in 1948 and 1952, respectively. He was a staff physicist at Lawrence Livermore Laboratory for twelve years and then president of New Mexico Institute of Mining and Technology for ten years, where he remains an Adjunct Professor. In 1976, he joined the Theoretical Division at the Laboratory and in 1980 became leader of the Theoretical Astrophysics Group. His research interests include nuclear physics, astrophysics, plasma physics, atmospheric physics, inertial fusion, geotectonic engineering, and the epidemiology of AIDS. Colgate has been involved in nuclear weapons testing, has participated in negotiations on nuclear testing for the U.S. State Department, and has had a leading role in magnetic fusion. His early work on supernovae led to the understanding of early neutrino emission from neutron stars—since then confirmed by SN1987A. In 1981, Colgate was named a Senior Fellow at the Laboratory and is an active member of the U.S. National Academy of Sciences.

Willy Benz is a professor of astronomy and planetary sciences at the University of Arizona, where he holds a joint appointment with Steward Observatory and the Lunar and Planetary Laboratory. He studied physics and earned a Ph.D. in astrophysics from Geneva University, Switzerland. Benz came to the United States in 1984 and joined the Laboratory as a postdoctoral fellow. In 1986, he moved to Harvard University as an assistant professor before joining the ranks of the faculty of the University of Arizona in 1991. Even though Benz left Los Alamos more than a decade ago, he has pursued scientific collaborations with several groups at the Laboratory as well as served as an occasional consultant. His research interests lie in the understanding of violent astrophysical phenomena ranging from dynamical fracture of brittle solids, collisions between planets and between stars, to supernova explosions. Benz has also been involved in investigating the collisional evolution of the early solar system, the origin of the moon, and more recently, the role of large-scale convection in supernova explosions.

Chris Fryer received his B.A. in mathematics and astrophysics at the University of California at Berkeley under the tutelage of Alex Filippenko. Unbeknownst to Fryer, that association sealed his fate. Almost immediately upon his arrival at the University of Arizona to pursue graduate work, he began working on supernovae with Willy Benz, who erroneously assumed that any student of Alex Filippenko's must necessarily want to work on such topics. Working on supernova and neutron-star projects naturally led to the acquisition of an additional advisor, Stirling Colgate. It also led to many "character-building" summers at Los Alamos. Fryer received his Ph.D. early in

1997 and is currently a postdoctoral fellow at the University of California at Santa Cruz, where he is working with Stan Woosley on supernovae, gamma-ray bursts, and neutron stars.

Dark Matter and Massive Neutrinos

Salman Habib

The Standard Model of particle physics does not include a natural mechanism to give mass to a neutrino. Neither does it provide a reason to exclude this possibility. Unlike photons and gravitons, which are constrained to be exactly massless because of gauge invariance, no such restriction applies to the neutrino. Determining neutrino masses has been a long-standing experimental challenge that, despite concerted efforts, has proved rather difficult. To date, there is no direct evidence for neutrino mass, though upper limits of $m_{\nu_e} < 15$ electron volts[1] (eV), $m_{\nu_\mu} < 170$ kilo-electron-volts (keV), and $m_{\nu_\tau} < 24$ million electron volts (MeV) have been placed.

It is natural to speculate what the impact on physics would be if neutrinos were in fact massive. As far as our everyday world is concerned, there would be almost no effect at all: nuclei would still undergo beta decay, elements would still transmute, and stars would still boil inside and explode because of neutrino heating. Solar and atmospheric neutrinos would still be missing, although physicists would be fairly certain as to where they went.

Turning to the Universe, however, massive neutrinos could effect a radical transformation. Next to the ubiquitous photons that compose the cosmic microwave background radiation (the radiation field that permeates the Universe), neutrinos are the second-most-abundant particle species. Were they to have even a small mass, it would lead to profound consequences for the evolution of the Universe. In this article, we explore the possible impact that neutrinos with mass would have on three central issues in modern cosmology: the dynamics of the Universe, structure formation, and dark matter.

Cosmology is the science of the evolution and structure of the Universe. The concerns of cosmology include the

birth of the Universe, its present age, and its ultimate fate. Some of the most pressing questions of current interest relate to the material make-up of the Universe: How much mass is present? What is it made of? How is mass distributed in space and how did it get there? A massive neutrino might well play a key role in the resolution of these puzzles.

According to the accepted theoretical paradigm in cosmology—the Big Bang—the Universe began as a hot, dense plasma that was isotropic and homogeneous to a very high degree. Fifteen billion or so years later, however, it is quite inhomogeneous (except on very large scales). Today the Universe is filled with galaxies that are arranged in clusters and sheets that surround vast pockets of space. Cosmologists attempting to understand structure formation must confront this puzzle: how did density fluctuations originate in the early Universe, and how did these small inhomogeneities lead to the distribution of mass that is currently observed?

Running parallel to the questions surrounding structure formation is the enigma of dark matter. After many years of observations, astronomers and cosmologists have been forced to a curious conclusion: the Universe appears to be dominated by an unseen form of matter whose precise nature is unknown.

For decades, it has been accepted that the luminous matter visible to the astronomer's telescope—stars, dust, gas clouds, bright galaxies, even black holes—constitutes but a tiny fraction of the total mass of the Universe. The phrase "luminous matter" refers to any matter that emits, directly or indirectly, electromagnetic radiation (from radio waves to gamma rays) that can be detected on earth. It is in this sense that large black holes can be considered luminous, for they advertise their presence by x-rays that are emitted when material falls into the hole.

Dark matter is the unseen mass of the Universe. It is the antithesis of

luminous matter, for it does not emit any detectable radiation, and its presence can be inferred only indirectly from the way it interacts with luminous matter. The three key questions relating to dark matter are what is it made of, how much is there, and how is it distributed?

Because it cannot be seen, we can only speculate as to what makes up dark matter. Many astronomers would argue that dark matter is simply stuff from the Universe's graveyard: brown dwarfs, dead stars, sparse gas clouds that never coalesced, even entire galaxies with low surface brightness. If this belief is true, dark matter would be ordinary *baryonic* matter—that is, matter composed of protons, neutrons, and electrons—that just fails to be detected.

Many theorists are convinced, however, that there is an exotic, *nonbaryonic* form of dark matter and that there is a lot more of it than ordinary matter. They hypothesize that the nonbaryonic dark matter is composed of particles that were created during the early, hot phase of the Universe but that still exist today. It is within this realm that the massive neutrino resides. While there are other plausible candidates for dark matter particles, such as axions and supersymmetric neutralinos, the neutrino is unique in that it is known to exist.

Because of improved observational capabilities, the last decade has seen a remarkable renaissance in astrophysics and cosmology. Telescopes such as the Keck and the Hubble Space Telescope, satellite experiments such as RELICT and the Cosmic Background Explorer (COBE), and large-scale redshift surveys such as CfA (conducted by the Harvard-Smithsonian Center for Astrophysics) and the Las Campanas Redshift Survey have changed the face of observational cosmology. As shown later on, the better quality of present-day data already allows us to rule out several plausible hypotheses concerning dark matter and structure formation.

In the coming decade, it is expected that data from new satellite missions that will measure the microwave background with unprecedented precision,

---

[1]This is a conservative upper limit on the electron neutrino mass. See R. M. Barnett et al., *Physical Review D* **54**, 1 (1996) for a complete discussion of mass limits.

combined with new, high-statistics redshift surveys that will probe the large-scale structure of the Universe, will finally lead to a cohesive picture of the Universe on large-distance scales.

Although the primary purpose of this article is to explain current theories about dark matter, structure formation, and the dynamics of the Universe, we do so with a word of caution. At present, the situation in cosmology is somewhat chaotic. Theorists are scrambling to keep pace with observational data, much of which is not simple to interpret and may contain significant systematic errors. Some of the observations discussed below are very recent and their validity may not survive over time, but given the state of cosmology today, they are all we have to work with at the moment.

Dark Matter: A Historical Problem

Astronomers tend to be a cautious lot, and with good reason. Ever since John Adams (in 1845) and Urbain-Jean-Joseph Le Verrier (independently in 1846) inferred the existence of the planet Neptune (through its gravitational effects on the orbit of Uranus), astronomers have appreciated that matter is often invisible to their telescopes.

It was therefore only a minor problem when Jacobus Kapteyn and James Jeans in 1922, and then Jan Oort a decade later, deduced that our galaxy, the Milky Way, might contain at least twice as much mass as could be accounted for by luminous matter alone. The missing mass would surely be found once more precise observations were made and the systematic errors identified and taken into account. Three-quarters of a century later, not only is the mass still "missing," but the fraction of galactic matter thought to be dark has increased.

The root of all such deductions is that mass is the sole source of the gravitational field, and gravity is the only force

we know of that binds a galaxy together and creates structure on cosmic scales. If a galaxy as a whole rotates, only the force of gravity can prevent it from flying apart. Measuring how fast a galaxy spins, therefore, gives an estimate of the strength of the galactic gravitational field, from which we can deduce the total mass needed to create that field.

The visible portion of a typical galaxy extends out about 15 kiloparsecs (kpc) from the galactic center.[2] Large clouds of atomic hydrogen, however, extend much farther, out to about 25 kpc. Measurements reveal that these clouds are in very high velocity orbits about the galactic center. A gravitational field strong enough to hold onto the hydrogen reaches well beyond the farthest stars in the galaxy.

Because the strength of gravity decreases inversely with the square of the distance from the source of the field, we are forced to conclude that the visible stars cannot be the *dominant* source of a galaxy's gravitational field. Rather, a "dark halo" of unseen matter must exist beyond the luminous core and must constitute the bulk of a galaxy's mass. This important deduction was first made by Kenneth Freeman in 1970.

How large is the halo? That is difficult to determine, because the hydrogen only extends so far, and until recently, there was no way to map out the gravitational field at arbitrarily large galactic radii. In fact, there does not exist a single spiral galaxy with a well-characterized halo or a well-determined mass! If we assume a "standard" galactic halo about 50 kpc in radius, then the total mass of a galaxy is roughly 10^{11} solar masses, or about a factor of 10 larger than estimates of the mass of the luminous core plus the hydrogen. Thus, by observing the dynamics of a galaxy, we conclude that roughly 90 percent of its structure consists of dark matter.

Current estimates, however, suggest that dark halos could go out much farther, to roughly 200 kpc. If these results hold up, then the amount of dark matter in galaxies could be

several times greater than the earlier dynamical estimates.

Galaxies are not the only type of structure in the Universe. They often group together to form clusters, which may contain hundreds to thousands of galaxies and span a distance of a few megaparsecs. Similar to the stars in a single galaxy, the galaxies in a cluster are all bound together by a common gravitational field. By examining the dynamics of constituent galaxies, we can estimate the total mass of the cluster. This technique was first applied to rich clusters by Fritz Zwicky back in 1933.

The Coma cluster of galaxies is an oft-cited example. The dispersion of random galactic velocities about the mean in this cluster is roughly 900 kilometers per second (km/s). Such high velocities demand that the cluster mass be about 5×10^{15} solar masses in order for the system to be gravitationally bound. However, the total luminosity of the cluster is only 3×10^{13} solar luminosities. This would give a ratio of the total mass to luminous mass (M/L) of close to 200 to 1, an estimate that is consistent with determinations from other rich clusters. Because the presence of x-ray-emitting gas increases the luminous mass by roughly a factor of 2, M/L is really more like 100 to 1. Thus, approximately 99 percent of the mass of a cluster comes from dark matter, a value that is roughly consistent with recent estimates of the amount of dark matter bound up in individual galaxies.

The data implies that the M/L ratio begins to approach a maximum at distances of hundreds of kiloparsecs to a megaparsec, which suggests that the distribution of mass in a cluster has resulted from a redistribution of the mass in individual galaxies. Although that deduction is pleasing, it runs counter to a prejudice held by many theorists that

---

[2] A parsec is equal to 3.258 light-years, or about 3×10^{13} kilometers. In general, the nearest stars are on the order of a few parsecs from earth, the diameter of a spiral galaxy spans tens of thousands of parsecs (kiloparsecs, or kpc), and distances between galaxies are about a million parsecs (megaparsecs, or Mpc).
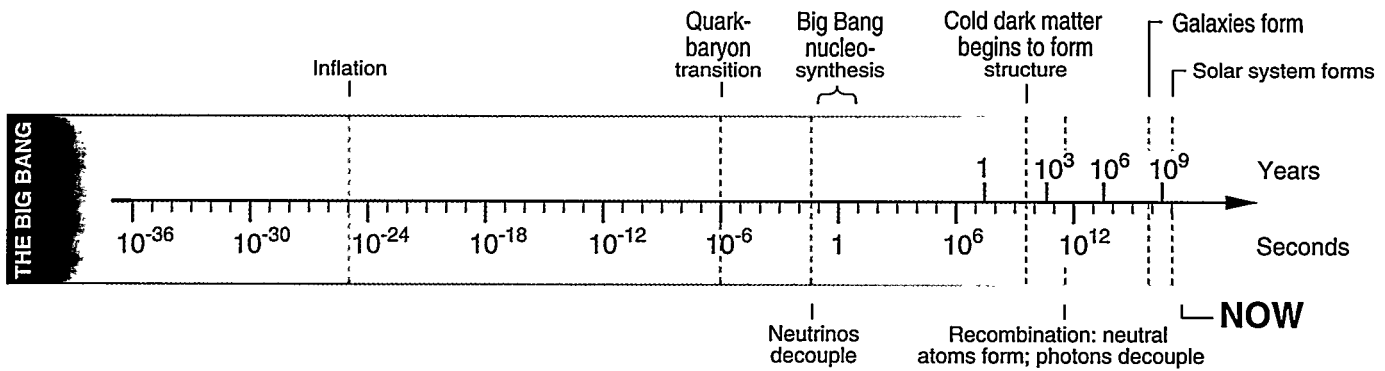
Figure 1. The Universe's Time Line

The Big Bang model allows cosmologists to order events in the evolution of the Universe. This figure plots time on a logarithmic scale. Although cold dark matter begins to form structures within the first 100 years or so of the Universe's history, those structures do not evolve into galaxies (or clusters of galaxies) until many millions of years later.

M/L should continue to increase with increasing distance scales. These theorists claim that a Universe consisting of only 99 percent dark matter is still too light and that there is more mass in the Universe than implied by the observations discussed so far. To explore their reasons, we must make a small digression into Big Bang cosmology.

The Big Bang: Dark Matter and the Dynamics of the Universe

One of the seminal discoveries in the history of science is Edwin Hubble's observation in 1929 that galaxies are receding from each other at a velocity v that is proportional to their distance l:

$$v = H(t) \, l \ .$$

The constant of proportionality, $H(t)$ (known as the Hubble constant), is actually a function of time. It is a difficult parameter to measure, but most cosmologists agree that its current value, $H(t_0) = H_0$, lies in the range of 55–75 km/s-Mpc. The uncertainty in the value of H_0 is contained in the parameter h, defined as $H_0 = 100 \, h$ km/s-Mpc.

Hubble's finding agreed with the velocity versus distance relationship predicted by Albert Einstein's general theory of relativity. The expanding Universe was therefore taken to be

strong evidence that general relativity correctly describes the dynamics of the Universe. Starting with present-day data, if the equations of general relativity are run backwards in time, the Universe becomes increasingly hotter and denser until the initial singularity, or a state of infinite density, is finally encountered. This is the moment of the Big Bang. If we run the clock forward from this moment (and use general relativity, particle and nuclear physics, electrodynamics, and thermodynamics to govern the interactions of matter, radiation, and geometry), we can construct a time line that orders the evolution of the Universe (see Figure 1).

The Big Bang model holds that the Universe began in a state of infinite density and temperature, followed by rapid expansion and cooling. About 10^{-30} seconds after its birth, quarks, leptons, and gauge bosons precipitated out much like ice crystals in a cooling pool of water. (Quarks and leptons are discussed in the primer, "The Oscillating Neutrino," on page 28.) Within a few microseconds, protons and neutrons formed from the quarks, and within about one second, the synthesis of primordial nuclei—hydrogen, helium, and trace amounts of lithium—began.

Primordial nucleosynthesis was completed by the time the Universe was about three minutes old, but the Universe was still too hot for the nuclei to capture electrons and form neutral

atoms. The Universe was filled with charged matter that continually scattered a background radiation field of energetic photons. Radiation and matter were in thermal equilibrium. Ten thousand years after the Big Bang, however, the expanding Universe had cooled to the point that atoms formed. This epoch, termed recombination, marks the time that radiation and matter decoupled. The Universe essentially became transparent to electromagnetic radiation, and radiation and matter began to follow separate evolutionary paths.

The primordial radiation field still permeates the Universe today and is essentially unchanged since the time of recombination. Because of the expansion of the Universe, the field has lost energy, and since a photon's energy is proportional to frequency, the radiation has shifted down into the microwave band. The existence of cosmic microwave background radiation (CMBR) was predicted by George Gamow, Ralph Alpher, and Robert Herman in 1948, and it was finally detected by Arno Penzias and Robert Wilson in 1965.

The Universe is still expanding today and literally creating its own space. What, however, determines whether it will do so forever, or will eventually deflate? The crucial parameter turns out to be the average mass density of the Universe, ρ. Below a certain critical mass density, ρ_c, the Universe is "open" and will forever

continue to expand. Above that critical density, the Universe is "closed" and will eventually recollapse into itself. At exactly the critical density, the Universe is said to be "flat." It is still infinite in space and time, but its rate of expansion asymptotically approaches zero. The critical density is easily derived; at the present time, it has the value

$$\rho_c = \frac{3H_0^2}{8\pi G} \quad ,$$

where G is Newton's gravitational constant. Today,

$$\rho_c \sim 4 \times 10^{-30} \text{ g/cm}^3 \quad ,$$

or a few hydrogen atoms per cubic meter. (We have assumed a value of $H_0 \approx 65$ km/s-Mpc.)

The critical, or closure, density provides a natural base line with which to compare observed mass densities. Defining a parameter Ω as the ratio of any density, ρ, to the critical density,

$$\Omega = \frac{\rho}{\rho_c} \quad ,$$

we have that at the critical density, $\Omega = 1$. The observed luminous, baryonic matter leads to

$$\Omega_{\text{luminous}} = \frac{\rho_{\text{luminous}}}{\rho_c}$$
$$\sim 0.003,$$

a mere 0.3 percent of the closure density, whereas measurements of clusters produce values for Ω_{cluster} within the range of 0.1 to 0.3.

Most theorists, however, believe that the Universe is at or extremely close to the critical density (in spite of the apparent conflict with current observations). The basis for their belief lies in the resolution of what is called the "flatness" problem in cosmology.

At birth, the Universe is postulated to be infinitely dense (the initial singularity). For generic initial conditions, once the Universe begins to expand, gravity under most circumstances causes it to recollapse instantly on

itself or to expand at such an enormous rate that no structures could ever form. This is because the natural time scale in general relativity is the Planck time, which is only about 10^{-43} seconds! The fact that the Universe has existed for 10^{60} Planck times cannot be explained without very special initial conditions (or entirely new physics). Only if the Universe started exquisitely close to the critical density could it have survived for such a long time.

The theory of inflation, which has become an almost essential piece of today's cosmology, was designed to deal with issues such as the flatness problem (also called the age problem). Inflation typically predicts deviations from the critical density on the order of only 1 part in 10^{60}. Inflation also accounts for the "horizon" problem, which stems from the observation that the cosmic microwave background is remarkably isotropic across the entire sky. This is a puzzle, because points in the sky separated in angle by more than roughly a degree have not been in causal contact since the Big Bang. Inflation provides a resolution to both problems by postulating a phase of rapid expansion of the Universe driven by a matter field called the inflaton. During inflation, the scale factor of the Universe grows by a factor of roughly 10^{43}. This growth occurs on a time scale as short as 10^{-32} seconds! In essence, inflation adds a long "history" to the Universe before the decoupling of radiation and matter, so that objects that appear to be acausally connected in the microwave sky actually interacted in the past. Finally, through quantum fluctuations of the inflaton field, inflation provides the Big Bang with a natural mechanism to generate primordial density perturbations. This is an important point that will be discussed in the section on structure formation.

Whether Ω is unity or on the order of a few tenths, it is apparent that the luminous fraction constitutes a very small proportion of the total mass of the Universe. The next natural question to ask is what is the composition of

this mysterious dark matter?

Again, we turn to Big Bang cosmology. One of the major achievements to emerge from that paradigm is the theory of Big Bang nucleosynthesis (BBNS). This theory describes how the two lightest baryons (protons and neutrons) could fuse together and form the light elements that are observed today in the cosmos. Protons (^{1}H) and neutrons first fused together to form deuterium. Fusion reactions involving deuterium then created tritium and helium-3, which were then used in further reactions to build helium-4 and trace amounts of lithium-7. These six elements (^{1}H, ^{2}H, ^{3}H, ^{3}He, ^{4}He, and ^{7}Li) are the only long-lived nuclei to be produced early on in the history of the Universe. (Heavier elements, from beryllium to uranium, were produced millions of years after the Big Bang by stellar nucleosynthesis and by supernova explosions.)

BBNS theory has only one free parameter, η, which is the primordial ratio of baryons to photons (or equivalently, the ratio of matter to radiation in the early Universe). Relative abundances of the primordial elements can be predicted as a function of η and compared with observations. This comparison leads to the estimate

$$\eta = \frac{\text{number of baryons}}{\text{number of photons}}$$
$$\sim 2.5 - 6 \times 10^{-10} \quad .$$

Now comes a remarkable bit of good fortune: an imprint of the primordial photon density still exists as the CMBR and has been measured with extreme precision by the COBE (satellite) and COBRA (rocket) experiments. Thus, from the estimates of η, we can estimate the primordial baryon density, that is, the total number of baryons that were produced during the Big Bang.

For many years, BBNS set a limit for the baryon density that was $\Omega_{\text{BBNS}} \sim 0.06$, a factor of 5 lower than the mass density estimate from clusters and only 6 percent of the value predicted by inflation. This was viewed as

184

unequivocal evidence for a nonbaryon-ic, massive particle, and several candidates were proposed: massive neutrinos, axions, neutralinos, quark nuggets, and primordial black holes.

However, there may yet be further surprises in store. It turns out that the estimate of η depends sensitively on the primordial abundance of deuterium. Deuterium absorption lines were recently measured in primordial intergalactic clouds illuminated by a background quasar. The conclusion was that previous estimates for deuterium abundance were too high; consequently, the value of η almost doubled, and Ω_{BBNS} could now be as large as 0.1. This value is not far from the preferred value of the mass density ascribed to clusters ($\Omega_{cluster} \approx 0.3$).

Given the overall uncertainty of the various mass density measurements, it is dangerous to predict just how much of dark matter is nonbaryonic. However, this fraction is likely to be at least two-thirds of all dark matter ($\Omega_{BBNS} \sim 0.1$ and $\Omega_{cluster} \approx 0.3$), and it could be much higher if Ω eventually turns out to be unity. These results are summarized in Table I.

The Big Bang: Structure Formation

One of the striking features of the Universe today—as opposed to the early Universe—is its inhomogeneity. Like islands and archipelagos in some vast ocean, matter floating in space has condensed into stars, planets, gas clouds, galaxies, and galactic clusters. Even the clusters seem to be organized into larger structures, creating great walls and sheets of galaxies that surround enormous bubbles or voids of lower density. Observations indicate that the Universe is "lumpy" on distance scales up to several tens of megaparsecs.

In earlier redshift surveys such as the CfA, there was strong inhomogeneity on the largest scales probed (~ 50 h^{-1} Mpc). (Although this distance

Table I. Comparison of Mass Densities

	Observation	Theory
$\Omega_{luminous}$	0.003	—
Ω_{galaxy}	0.02–0.1	—
$\Omega_{cluster}$	0.1–0.3	—
$\Omega_{baryonic}$	—	0.01–0.1 (BBNS)
Ω_{total}	0.1–1	1 (inflation)

is on the order of 300 million light-years, the survey probed but a tiny fraction of the observable Universe, which is estimated to be about 3000 h^{-1} Mpc across.) However, much deeper surveys such as Las Campanas (~ 600 h^{-1} Mpc) provide evidence that on very large distance scales, the size of structures saturate and no longer increase. The Universe is apparently homogeneous on scales greater than about 100 h^{-1} Mpc (see Figure 2).

A major triumph of the standard Big Bang model has been the progress made in understanding structure formation as a result of the gravitational Jeans instability. It turns out that the evolution of small perturbations of a uniform background density can be studied in much the same way as the stability properties of an ordinary plasma. The Jeans instability comes about because gravity always attracts: above a certain wavelength, called the Jeans length, density fluctuations are unstable and grow exponentially. In an expanding Universe, this exponential growth is modified and slows down to a weak power law.

An important aspect of the Jeans instability is that it does not saturate at some finite value from nonlinear feedback, but rather increases in strength as the gravitational collapse proceeds. It stops increasing only when the structures formed have enough internal energy—for example, gas pressure in stars and kinetic energy in the solar system—

to be able to resist collapsing further.

Another subtlety that has to be taken into account is the growth of density perturbations in the presence of thermal radiation. In the early history of the Universe, when matter is in the form of an ionized plasma and the energy density in radiation is much greater than that of matter, there is a strong coupling between radiation and matter. The radiation field itself does not collapse, and it prevents matter from collapsing because of the strong coupling. Only perturbations on scales longer than the Jeans wavelength, given by

$$\lambda_J = v_s \sqrt{\frac{\pi}{\rho G}},$$

where v_s is the velocity of sound, continue to grow. Smaller-scale perturbations oscillate as damped acoustic waves. After recombination, the velocity of sound drops abruptly as the pressure support switches from radiation to neutral hydrogen. Consequently, density perturbations on much smaller scales can also begin to grow.

This picture of how initial density perturbations grow into structures is attractive, but it lacks a key ingredient: a source for the initial density perturbations that the Jeans instability would then amplify. The original Big Bang model does not have a physical mechanism to produce these perturbations. But the precise nature of the perturbation spectrum is very important, for it controls sensitively the types of
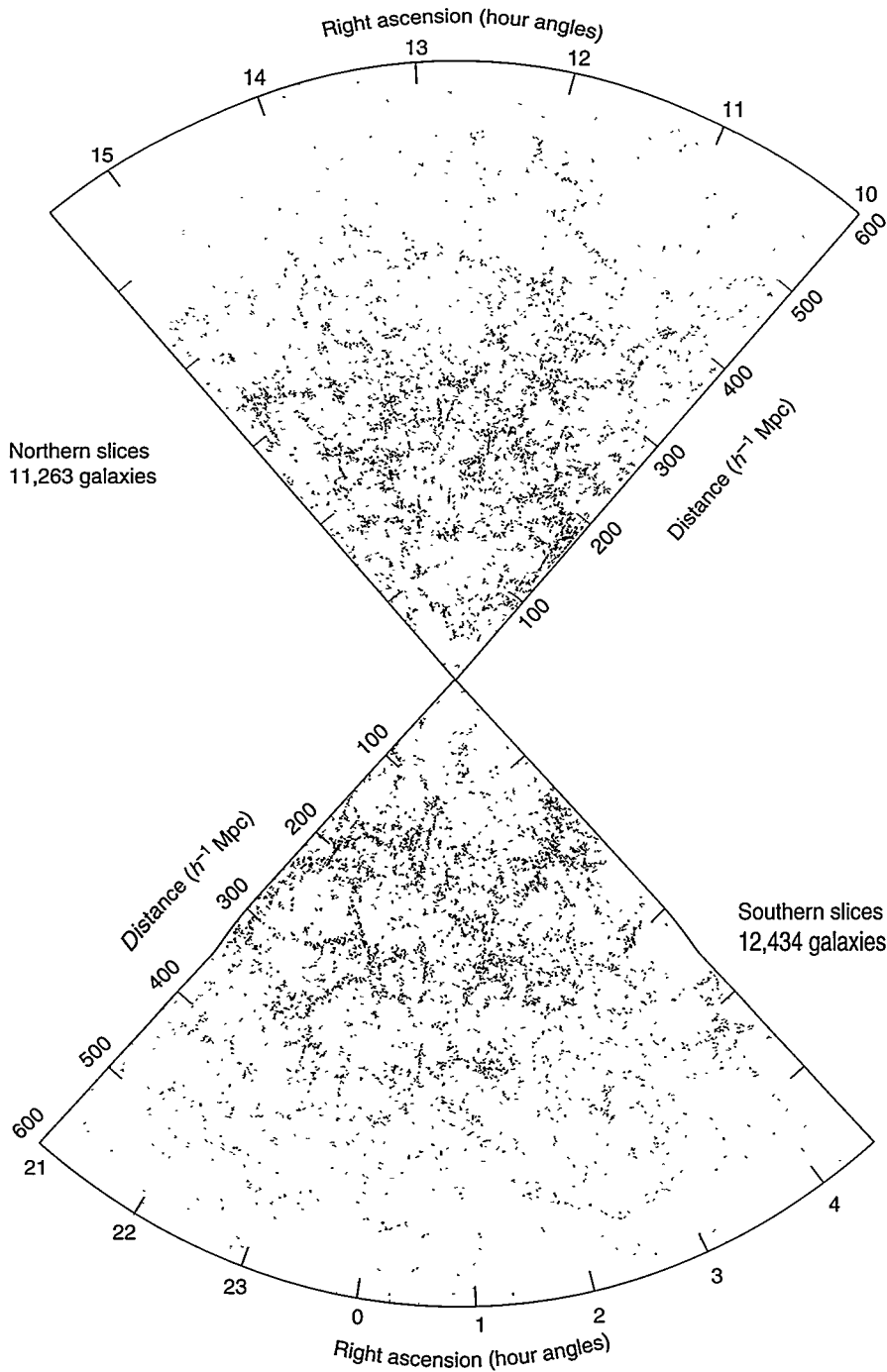
Figure 2. Result of the Las Campanas Redshift Survey

This map of over 23,000 galaxies extends to approximately 600 h^{-1} Mpc, or about one-fifth of the observable Universe. Galaxies brighter than 19th magnitude were counted in various "slices" of the sky. Each slice spanned about 90 to 120 degrees and was confined to a plane oriented at some angle (declination) with respect to the celestial equator. Data from three northern and three southern slices have been superimposed in this figure. The scale of the largest structures (the "voids" containing few galaxies) is roughly 50 h^{-1} Mpc, and there is no evidence for large-scale inhomogeneity on the scale of this survey. (The change in the galactic density beyond about 400 h^{-1} Mpc is an artifact. At great distances, the survey detects only the brighter galaxies.)

For more information about Las Campanas, see http://manaslu.astro.utoronto.ca:80/~lin/lcrs.html.

structures that form at later times. At present, the Universe has evolved numerous complex, scale-dependent structures, and the simplest primordial spectrum of density fluctuations that could potentially lead to what is observed today is the one put forth by Edward Harrison (in 1970) and Yakov Zeldovich (independently in 1972).

The Harrison-Zeldovich spectrum was based on very general theoretical considerations and has been used as the initial density perturbation spectrum in most analytical studies and simulations that attempt to track the Big Bang. This spectrum implies that the amplitude of primordial fluctuations in the gravitational potential does not depend on the spatial scale and, for a critical-density matter-dominated Universe, is also independent of time.

Significantly, the Harrison-Zeldovich spectrum also emerges from inflation theory. Quantum fluctuations of the inflaton field that drives the inflationary expansion provide a natural source of density perturbations that follow a Harrison-Zeldovich spectrum. Thus, aside from solving the flatness and horizon problems, inflation builds into the Big Bang model a natural mechanism for generating initial density perturbations.

Given the observational constraints and a primordial density perturbation spectrum, the question is whether the Jeans instability successfully produces the large-scale structures that are observed today. One point to address is the normalization—that is, the absolute amplitude—of the primordial density fluctuations. Simply choosing a spectrum does not determine its absolute scale. The normalization needs to be determined by experiment, but how do we measure the size of density fluctuations that were present 15 billion years ago? Remarkably, a window to the past exists that allows us to do just that: measuring anisotropies in the CMBR temperature provides a virtual time-machine to determine the lumpiness of the very early Universe.

The discovery of the microwave

background was a stunning confirmation of the Big Bang, but detection of a temperature anisotropy in the field, or deviation from a perfectly uniform temperature, could have an even greater impact. Photons that make up the microwave field have been traveling unimpeded since the time of recombination. Because of intrinsic fluctuations in the temperature and gravitational potential of the Universe at the time the photons decoupled from matter, there is a very small anisotropy in the CMBR temperature observed today.

The anisotropy over large-distance scales was measured with very high precision by the COBE satellite, launched in 1989 (see Figure 3). (COBE's angular resolution of 7 degrees corresponds to several hundred megaparsecs.) The COBE results, along with those of other experiments that probed the microwave background at higher angular resolution, set the normalization of the Harrison-Zeldovich spectrum (see Figure 4) and impose constraints on any proposed spectrum of initial density perturbations. One important consequence of the CMBR observations is that the observed large-scale structure of the Universe cannot have formed in the presence of ordinary baryonic matter alone.

In purely baryonic matter models, the growth of initial perturbations occurs only after recombination. As stated earlier, before that time, growth is suppressed by pressure that arises when radiation scatters from free electrons (Thomson scattering), resulting in the effective prevention of growth of perturbations on scales smaller than $\sim 180\ h^{-1}$ Mpc. To produce the observed large-scale structure requires that the perturbations leave an anisotropy in the microwave background temperature of roughly 1 part in 10,000 on the scales probed by COBE. But the measured fluctuations were much smaller, deviating from pure uniformity by only 1 part in 100,000. The microwave background, therefore, informs us that there was insufficient time for structure formation in a purely baryon-dominated Universe.
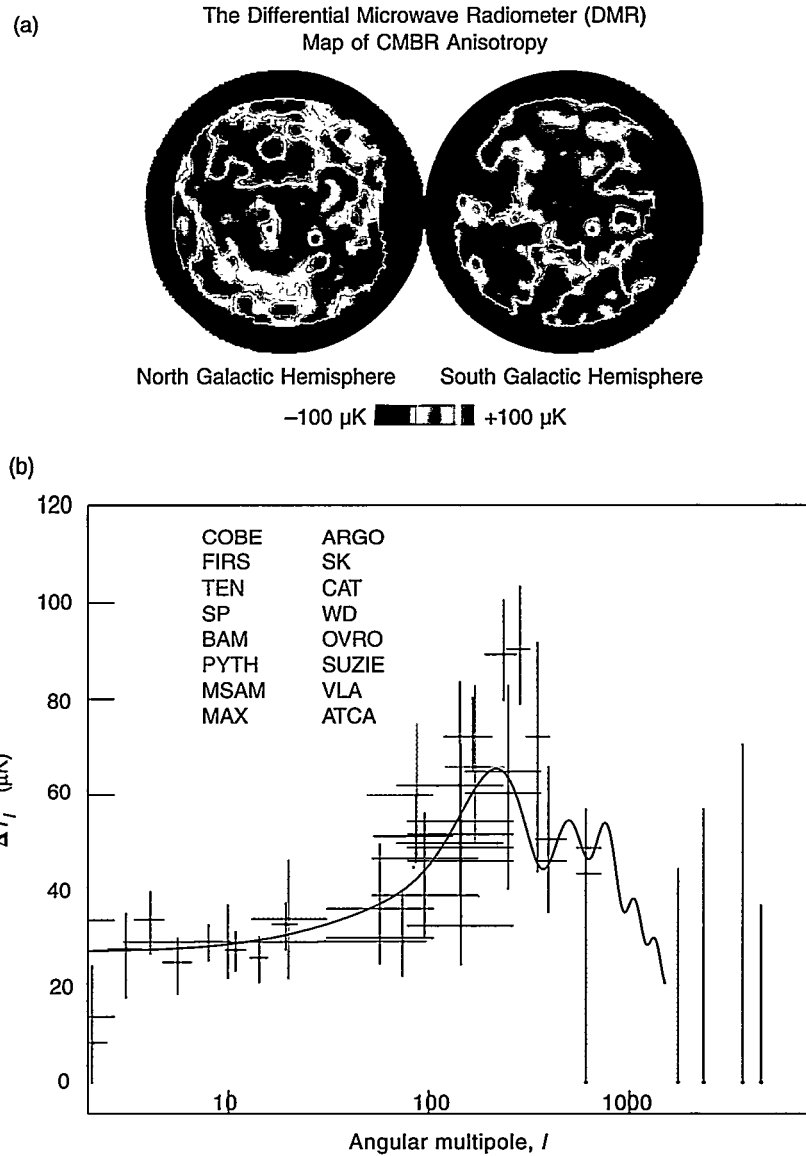

Figure 3. Temperature Fluctuations across the Microwave Sky

(a) The DMR experiment on the COBE satellite measured root-mean-squared (rms) temperature variations, $\Delta T/T$, in the CMBR to be on the order of 1 part in 10^5. (The average temperature of the background is 2.728 ± 0.004 kelvins.) The variations can be related to density fluctuations at the time of recombination that seeded the current large-scale structures seen in Figure 2. (The scale of the map shown in (a) is enormous. The largest structure of Figure 2 would easily fit within the smallest feature of the map.) (b) Data from 16 experiments that have measured the CMBR with varying degrees of angular resolution are shown in this figure of ΔT_l (the rms temperature fluctuations per logarithmic interval $\Delta l/l$) versus the angular multipole l (plotted on a log scale). The black curve is a theoretical prediction for the CMBR, based on a cold dark matter model (discussed later in the text.) At very large distance scales (small l), the anisotropy is determined by the primordial power spectrum and is predicted to be a flat line for the Harrison-Zeldovich spectrum. Values of $l < 80$ correspond to distance scales that were not causally connected at the time of recombination.
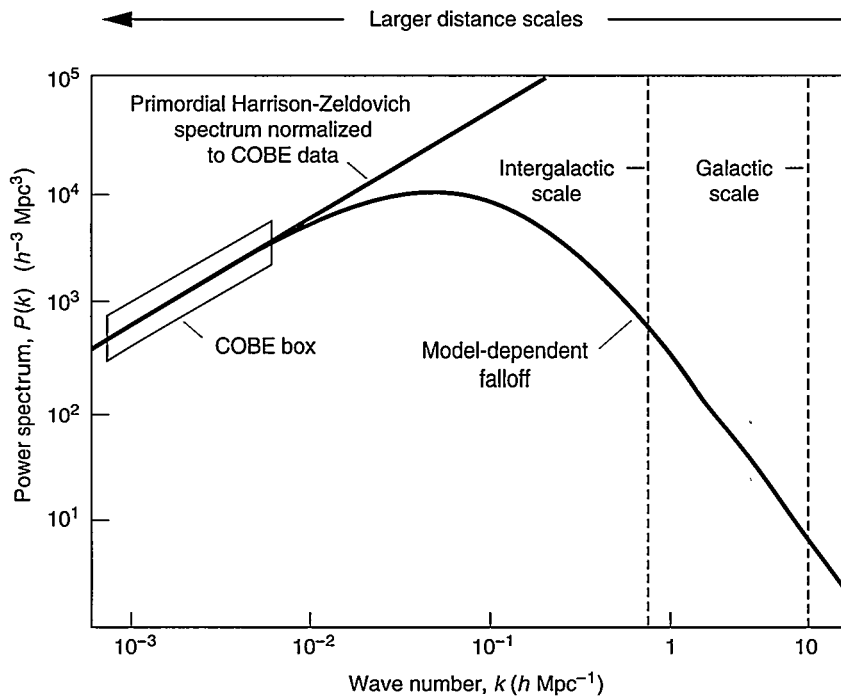
Figure 4. The Harrison-Zeldovich Spectrum and a Processed Spectrum

The primordial Harrison-Zeldovich power spectrum (shown in red) is plotted against wave number, *k*, on a log-log scale. (Notice that *k* varies inversely with distance.) Data from COBE (rectangular box, derived from the CMBR data at large angular multipoles) fixes the normalization, that is, the position of the spectrum in the vertical direction. The power spectrum that one compares with observations is a processed spectrum like the one shown in black. At low *k*, the spectrum is Harrison-Zeldovich, while the falloff at large *k* is model dependent. For any given model, the processed power spectrum is the Harrison-Zeldovich spectrum multiplied by a transfer function, *T(k)*, which incorporates the contribution of all relevant physical processes.

Nonbaryonic dark matter, however, does not couple to photons and thus does not suffer from collisional damping. Density perturbations can begin to grow well before recombination, as soon as matter-radiation equality is achieved. This allows the development of large density perturbations without violating the density constraints implied by the small anisotropy in the microwave background. Some form of nonbaryonic dark matter, therefore, appears to be necessary to explain the formation of structure in the Universe.

Despite the complete absence of direct experimental evidence for a nonbaryonic, dark matter particle, theorists have had no trouble in suggesting a plethora of possible candidates. Although their specific properties vary,

all candidates are bound by the common constraints that they have mass, not be made of quarks, and have neither strong nor electromagnetic interactions.

Structure Formation and Dark Matter

Of all the proposed dark matter candidates, massive neutrinos have always been the most natural: neutrinos are known to exist, and they were produced in very large numbers during the Big Bang. (Roughly a billion neutrinos were created for every baryon.) Since the mean density of the Universe has to satisfy observational constraints, there exists a calculable range of allowed neutrino masses. Assuming a single,

two-component neutrino species, the relevant formula is

$$m_\nu = 92\ \Omega_\nu\, h^2\ ,$$

where the neutrino mass m_ν is measured in electron volts. If we assume that the Universe is at critical density, this yields a neutrino mass of roughly 30 eV (assuming $h \approx 0.6$). With $\Omega_\nu = 0.2$, the mass range falls to several electron volts.

However, structure formation must also be considered. We would like to know in what way the evolution of structure depends on the type of nonbaryonic dark matter. This question leads to a simple hot or cold classification of dark matter, a classification based on the random velocities of dark matter particles at the moment they fall out of thermal equilibrium with the photon heat bath in the expanding Universe. Relativistic particle velocities are characteristic of hot dark matter, while cold dark matter particles are either very heavy, and hence nonrelativistic early on, or are created with essentially no random velocity.

In the case of massive neutrinos, which decouple from the rest of the Universe at a temperature of 1 MeV, masses on the order of tens of electron volts or less make them highly relativistic. They are therefore an excellent (and currently the only) candidate for hot dark matter. In a massive neutrino, hot dark matter model with adiabatic initial perturbations (radiation strongly coupled to matter), typically very large scale structures form first. The large bodies then break up to form objects at smaller scales: this is "top-down" structure formation.

Such a growth pattern results because small-scale structure in hot dark matter models is washed away by neutrino free-streaming, a collisionless effect analogous to Landau damping in plasmas. As long as density fluctuations are stable against the Jeans instability, collisionless particles such as neutrinos can stream out of the higher-density regions into the lower-density regions,
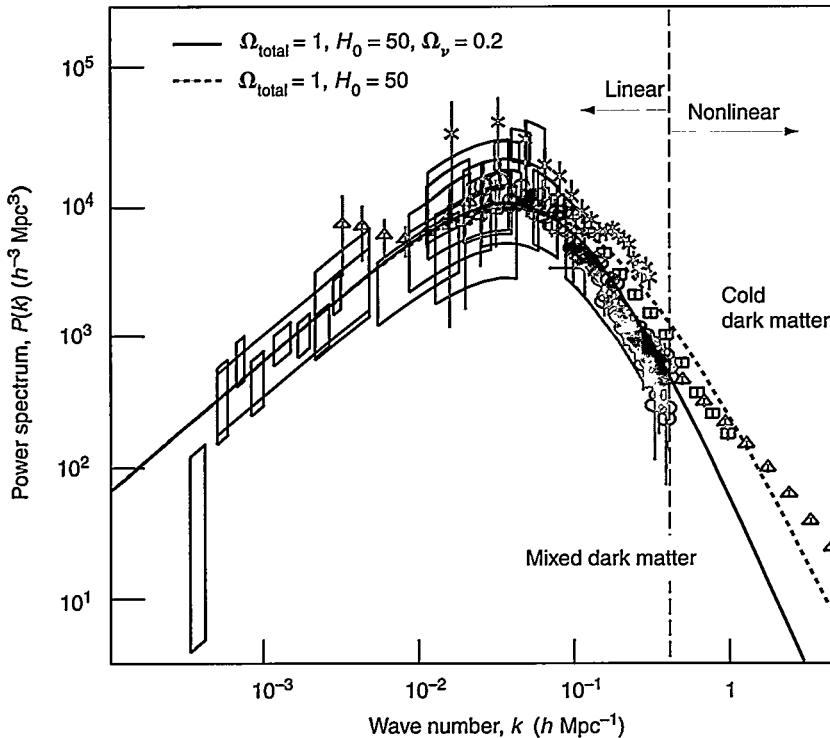
Figure 5. Large-Scale Structure Data and Dark Matter Power Spectra
Two theoretical, linear power spectra (best-fit mixed dark matter and standard cold dark matter) are shown superimposed on observational data. The black boxes are reconstructed in a model-dependent way from the measurements of the CMBR data shown in Figure 4 and are given here for mixed dark matter. (The box height reflects a 1-sigma confidence level. The boxes differ slightly near the peak of $P(k)$ if a cold dark matter model is assumed.) Observations from matter surveys are shown in light grey. For $k > 0.3$ h Mpc^{-1}, the data is measuring nonlinear structure, beyond which point it cannot be directly compared with the linear theoretical power spectra. The over-production of small-scale structure by cold dark matter models is best seen in the region around k of 10^{-1} h Mpc^{-1}, where mixed dark matter is very successful.
Figure courtesy of E. Gawiser and J. Silk, CMB Theory group, UC Berkeley.

thereby smoothing inhomogeneities. Regions separated by distances larger than the neutrino free-streaming length survive this smoothing, so that on larger scales, differences in density are maintained and can grow.

Unfortunately, hot dark matter models have been shown to be incompatible with observations. Assuming the Harrison-Zeldovich spectrum predicted by inflation and normalized to COBE, numerical simulations have shown that, owing to suppression of small-scale fluctuations by neutrino free-streaming, cluster and galaxy formation occurs far too close to the present time. Another difficulty is that the large coherence

length of neutrinos makes it difficult for them to explain the localization of dark matter in individual galaxies. Thus, any model containing only adiabatic fluctuations and massive neutrinos has been ruled out.

Cold dark matter is composed of particles that are massive enough to become nonrelativistic shortly after their birth (or that are born with no random velocity). Although there is no experimental evidence for such particles, the supersymmetric candidates and the axion (Peccei-Quinn symmetry)—particles that are associated with central ideas in high-energy physics—have been suggested as possible cold dark

matter candidates. Because it is nonrelativistic, cold dark matter readily clumps together, and structure formation typically proceeds in a "bottom-up" manner. Galaxies form first, which then get grouped into clusters and superclusters, possibly in a hierarchical way.

However, predictions from cold dark matter models also disagree with observations. With the Harrison-Zeldovich assumption for the primordial density perturbations plus the COBE anisotropy constraints, we find that cold dark matter tends to produce too much structure at small scales.

The complementary nature of structure formation theories—massive neutrinos generate too little structure at small scales, while cold dark matter overproduces it—has naturally led to a model enjoying some popularity at present: mixed dark matter. Both kinds of dark matter are assumed to exist. Adding some massive neutrinos to a predominantly cold dark matter model tends to reduce the overproduction of small-scale structure characteristic of pure cold dark matter because of the neutrino free-streaming. It has been shown that in a critical-density Universe, if we choose a cold to hot dark matter ratio of 5 to 1 in mean mass density, then a mixed dark matter model might be viable and escape the serious problems of both the hot and cold dark matter models (see Figure 5). (There are indications of problems in forming structure at early enough times, but it is difficult to tell how serious these problems really are.) For such a mixed dark matter model to work, the neutrino mass has to be in the range of several electron volts, and this mass range is compatible with results from the liquid scintillator neutrino detector (LSND) experiment.

But mixed dark matter models also face tight constraints. One problem is that even a small admixture of hot dark matter reduces structure formation on small scales to the extent that a fairly large amount of cold dark matter is needed to compensate. It turns out that even for $m_\nu \sim 2$ eV, corresponding to
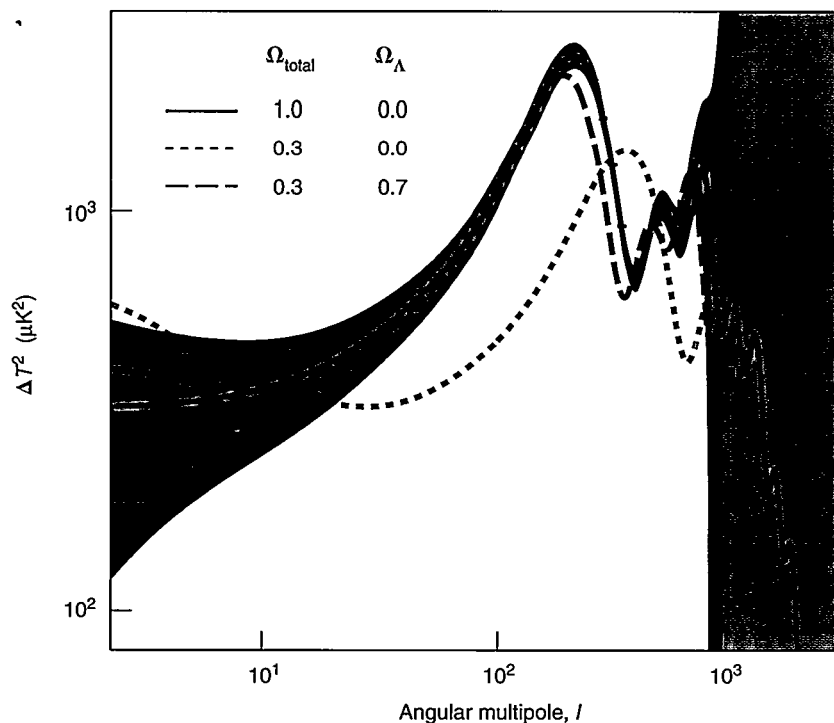
Figure 6. MAP and the Future

Temperature fluctuation squared is plotted to show how the MAP mission can distinguish between different theoretical models, here cold dark matter (dark blue curve), an open model with $\Omega = 0.3$ (dashed red curve), and a model with a nonzero cosmological constant Ω_Λ (dashed green curve). The rectangles that follow the blue curve are projected error bars, while the solid-blue envelope represents the so-called "cosmic variance," a finite sampling effect due to the fact that the observation is restricted to only a fraction of the Universe. Thus MAP will provide useful information for multipoles up to $l \sim 1000$ and is dominated by cosmic variance up to $l \sim 600$. If this figure corresponded to real data, only the cold dark matter model would be viable.

Figure courtesy of Wayne Hu; available at http://www.sns.ias.edu/~whu/physics/physics.html.

$\Omega_\nu \sim 0.1$, a large value of Ω_{total} is favored. The contribution from the cold dark matter (Ω_{CDM}) must be at least greater than 0.3, which is already at the upper range of observational limits. If Ω_{total} turns out to be low, mixed dark matter models would be strongly disfavored.

Another alternative model for structure formation is based on decaying heavy neutrinos. The overproduction of small-scale structure in cold dark matter models can be mitigated with either of two strategies, either by reducing the density of cold dark matter particles as in mixed dark matter models or by increasing the energy in radiation. The latter can be accomplished with unstable cold dark matter particles that decay.

Decaying-neutrino models are characterized by two parameters: the mass of the neutrino, m_ν, and its lifetime, t_d. In the early Universe, when the thermal energy of the neutrinos is much greater than m_ν, they behave as essentially massless particles. At later times, when the temperature falls, the energy density in the Universe can be dominated by these species, after which time they decay and release their rest energy into relativistic particles. After this point, the evolution is similar to that of a cold dark matter model except for the additional energy density due to relativistic particles. This is what allows for the tailoring of the perturbation power spectrum in order

for the model to be observationally viable. The acceptable mass range is $m_\nu \geq 10$ keV, while decaying neutrinos with $m_\nu \leq 50$ eV are ruled out (for all values of t_d).

So far, our discussion of both hot and cold dark matter models has assumed the Harrison-Zeldovich form for the primordial spectrum with adiabatic perturbations. A natural question is whether the situation is any different when other types of perturbations are considered. An alternative to inflation in this respect comes from theories in which the initial density perturbations are seeded through the formation of topological defects in early Universe phase transitions.

In these theories, topological defects such as cosmic strings give rise to perturbations either through the formation of wakes of overdensity as they move through the Universe or through the accretion of matter onto string loops. Calculations with these models are much more difficult than with inflationary models. Until recently, cosmic string/hot dark matter models were viable candidates for large-scale structure formation; unfortunately, topological defect models have now been ruled out. For a given anisotropy in the CMBR temperature, the corresponding amplitude in density fluctuations is several times too low to explain structure formation. In addition, the predicted small-angular-scale CMBR anisotropies are in conflict with present ground-based and balloon-borne observations.

Outlook

Although we have presented an up-to-date summary of dark matter and its relationship to structure formation, it should be noted that the outcome of models of structure formation and CMBR anisotropy depends critically on the values of cosmological parameters such as Ω_{total} and the Hubble constant H_0. It is important to pin down their values to within a few percent. At

190

present, the constraints on these parameters from CMBR anisotropy measurements are quite weak.

But the future is full of promise. The constraints on Ω_{total} and H_0 are expected to improve dramatically with the next generation of satellite observations. The Microwave Anisotropy Probe (MAP) is scheduled to fly in 2001, followed several years later by the PLANCK satellite. In addition, deep, high-statistics redshift surveys of galaxies are expected to yield data within the next several years. The 2 Degree Field (2dF) and the Sloan Digital Sky Survey (SDSS) will go out about the same distance as Las Campanas (roughly 600 h^{-1} Mpc), but they will survey many more galaxies: a quarter of a million for 2dF and a million for SDSS, compared with roughly 25,000 for Las Campanas.

Analysis of the new CMBR data combined with the large-scale structure information from the redshift surveys will provide a very powerful discriminator between competing models of structure formation (see Figure 6). A value of $\Omega_{total} \ll 1$ would be unfavorable for models incorporating light neutrinos and would be difficult to reconcile with inflation. (If the matter density is less than critical, that is, $\Omega_{matter} < 1$, it is still possible to save $\Omega_{total} = 1$ by introducing a large cosmological constant, an alternative espoused by some theorists.) On the other hand, if standard inflation is vindicated and $\Omega_{matter} = \Omega_{total} = 1$, a light neutrino might be just what theory needs to satisfy the constraints imposed by structure formation. Even if this were the case, however, the neutrino would still not play a major role in dictating the dynamics of the Universe.

It is unlikely that the last word has been spoken on the cosmological consequences of a massive neutrino. Today, such a particle is not the favored dark matter candidate given our theories of initial conditions and structure formation. Just how good or bad these theories are will not be known until the next generation of

CMBR observations yield results. Can the massive neutrino regain the dark matter center stage? The turn of the millennium may bring us the answer to that question. ∎

Further Reading

Coles, P., and F. Lucchin. 1995. *Cosmology: The Origin and Evolution of Cosmic Structures.* New York: Wiley.

Hu, W., N. Sugiyama, and J. Silk. 1997. The Physics of Microwave Background Anisotropies. *Nature* 386: 37.

Mather, J. C., and J. Boslough. 1996. *The Very First Light: The True Inside Story of the Scientific Journey Back to the Dawn of the Universe.* New York: Basic Books.

Partridge, R. B. 1995. *The Cosmic Microwave Background Radiation.* New York: Cambridge University Press.

Peebles, P. J. E. 1993. *Principles of Physical Cosmology.* Princeton: Princeton University Press.

Primack, J. R. 1996. Dark Matter and Structure Formation in the Universe. In *Formation of Structure in the Universe: Proceedings of the Jerusalem Winter School 1996.* Edited by A. Dekel and J. P. Ostriker. New York: Cambridge University Press.

Sahni, V., and P. Coles. 1995. Approximation Methods for Nonlinear Gravitational Clustering. *Physics Reports* 262: 2.

Schaeffer, R., J. Silk, et al., Eds. 1996. *Les Houches Summer School Proceedings: Cosmology and Large Scale Structure,* Vol. 60. Amsterdam: North-Holland.

Silk, J. 1997. *A Short History of the Universe.* New York: W. H. Freeman and Co.
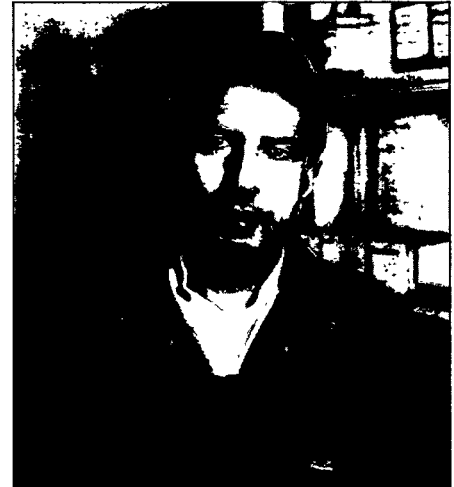
Salman Habib received his undergraduate degree in physics from the Indian Institute of Technology, Delhi, and his Ph.D. from the University of Maryland, where he was in the General Relativity and Cosmology group. After postdoctoral stints at the University of British Columbia and Los Alamos National Laboratory, Habib became a lstaff member of the Elementary Particles and Field Theory group at Los Alamos. His research interests cover many aspects of nonlinear dynamical systems, from nonequilibrium quantum field theory to the physics of charged-particle beams. Habib has a long-standing interest in the early Universe, especially the formation of large-scale structure and the associated problem of dark matter.