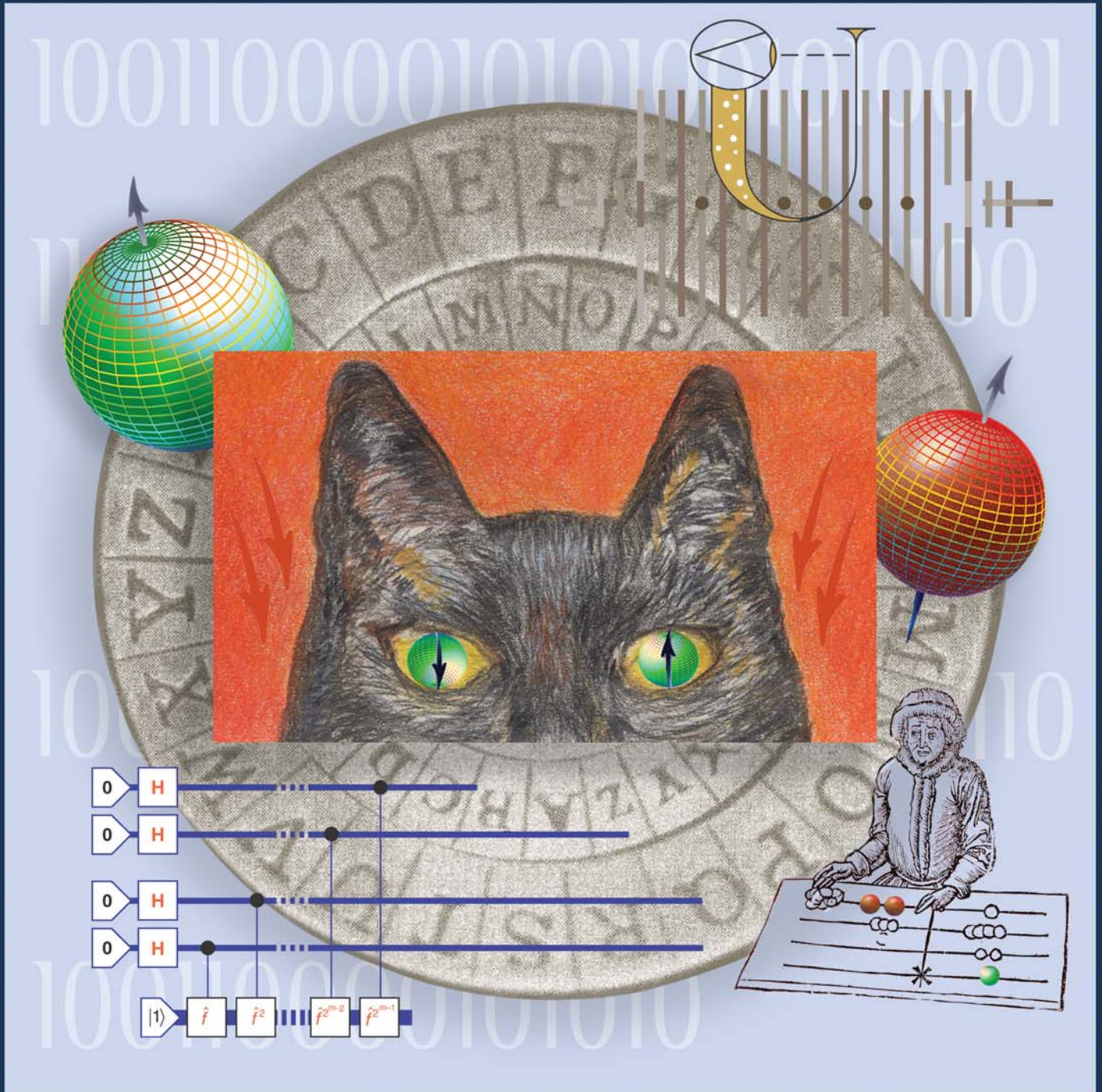


Los Alamos Science

LOS ALAMOS NATIONAL LABORATORY



About the Cover

A strange-looking cat welcomes you to the rich world of quantum research. Its sphere-like eyes represent the qubit, the quantum version of a classical bit. They point “up” and “down” in a quantum superposition, reminding us of Schrödinger’s famous “cat paradox.” A cat is trapped in a steel chamber with a “diabolical device”—a flask of cyanic acid attached to a Geiger counter containing a tiny bit of very long-lived radioactive material.



The decay of one atom will cause the contraption to shatter the flask and poison the cat. What is the quantum mechanical wave function of this system before we look inside? It is a strange superposition in which the cat is dead and alive at the same time! Clearly, Schrödinger warns us not to ascribe too much reality to the wave function.

But perspectives change. Almost seven decades later, scientists are manipulating quantum superpositions in ways that make them seem almost tangible. In computation and communication, single qubits and multiple qubits are presenting new opportunities.

One example is Peter Shor’s famous quantum factoring algorithm. The beginning of the relevant quantum computing network is shown at the lower left. Another example is quantum cryptography, the new wave for communicating secret keys and a remarkable departure from the stone cipher wheel of long ago seen in the background.

At upper right is John Wheeler’s drawing of the universe, a giant U, with the observer, a big eye, looking backward in time. The thin upper right end of the U represents the Big Bang, when it all started. Moving down, along the thin right leg, and up, along the thick left leg of the U, symbolically traces the evolution of the universe—from small to large. It is by observing single photons from the distant past that the early universe becomes part of our reality. In Wheeler’s view, our reality ultimately derives from measurement of individual quanta—“it from bit.”

(The drawing of the abacist at bottom right is used with permission from Cliché Bibliothèque Nationale de France, Paris. Permission for use of the confederate cypher wheel is from the Louis Kruh Collection.)

For past *Los Alamos Science* issues,
see our Web site at the following URL:
<http://www.lanl.gov/external/science/lascience/index.html>

Editor

Necia Grant Cooper

Managing Editor

Ileana G. Buican

Science Writer

Jay A. Schecker

Designer

Gloria E. Sharp

Illustrators

Andrea J. Kron

Chris D. Brigman

David R. Delano

Editorial Support

Faith J. Harp

Composition Support

Joy E. Baker

Wendy M. Burditt

Jeanne M. K. Bowles

Photographers

Richard C. Robinson

John A. Flower

Printing Coordination

Guadalupe D. Archuleta

Address mail to

Los Alamos Science

Mail Stop M711

Los Alamos National Laboratory

Los Alamos, NM 87545

lascience@lanl.gov

Fax: 505-665-4408

Tel: 505-667-1447

Information, Science, and Technology in a Quantum World

John Wheeler and Richard Feynman on Quantum Theory and Information	vi
Preface	x
About This Volume and Quantum Research at Los Alamos	1

Concepts in Quantum Information Science

Quantum Information Processing—A Hands-on Primer	2
<i>Emanuel Knill, Raymond Laflamme, Howard N. Barnum, Diego A. Dalvit, Jacek J. Dziarmaga, James E. Gubernatis, Leonid Gurvits, Gerardo Ortiz, Lorenza Viola, and Wojciech H. Zurek</i>	
Glossary	33
From Factoring to Phase Estimation—A Discussion of Shor’s Algorithm	38
<i>Emanuel Knill, Raymond Laflamme, Howard N. Barnum, Diego A. Dalvit, Jacek J. Dziarmaga, James E. Gubernatis, Leonid Gurvits, Gerardo Ortiz, Lorenza Viola, and Wojciech H. Zurek</i>	
20 Questions, Quantum Computers, and Cryptography	46
<i>Mark Ettinger</i>	
Quantum State Entanglement—Creation, Characterization, and Application	52
<i>Daniel F. V. James and Paul G. Kwiat</i>	
A New Face for Cryptography	68
<i>Jane E. Nordholt and Richard J. Hughes</i>	

Quantum Science

Decoherence and the Transition from Quantum to Classical— <i>Revisited</i>	86
<i>Wojciech H. Zurek</i>	
The Emergence of Classical Dynamics in a Quantum World	110
<i>Tanmoy Bhattacharya, Salman Habib, and Kurt Jacobs</i>	
Quantum Feedback Control—How Can We Control Quantum Systems without Disturbing Them?	126
<i>Salman Habib, Kurt Jacobs, and Hideo Mabuchi</i>	
Atom-Trap BECs—A New Laboratory for Studying Superfluidity, Quantum Fluctuations, and Other Quantum Phenomena	136
<i>Eddy M. E. Timmermans</i>	

Schrödinger Cats in Atom-Trap BECs	166
<i>Diego A. R. Dalvit and Jacek Dziarmaga</i>	
Experiments on Cold Trapped Atoms	168
<i>David J. Vieira and Xinxin Zhao</i>	
Quantum Information with Trapped Strontium Ions	178
<i>Dana J. Berkeland</i>	
Theory of Single-Spin Detection with a Scanning Tunneling Microscope	184
<i>Alexander V. Balatsky and Ivar Martin</i>	

Quantum Computation

Introduction to Quantum Error Correction	188
<i>Emanuel Knill, Raymond Laflamme, Alexei Ashikhmin, Howard N. Barnum, Lorenza Viola, and Wojciech H. Zurek</i>	
NMR and Quantum Information Processing	226
<i>Raymond Laflamme, Emanuel Knill, David G. Cory, Evan M. Fortunato, Timothy F. Havel, Cesar Miquel, Rudy Martinez, Camille J. Negrevergne, Gerardo Ortiz, Marco A. Pravia, Yehuda Sharf, Suddhasattwa Sinha, Rolanda Somma, and Lorenza Viola</i>	
Realizing a Noiseless Subsystem in an NMR Quantum Information Processor	260
<i>Lorenza Viola and Evan M. Fortunato</i>	
Ion-Trap Quantum Computation	264
<i>Michael H. Holzschneider</i>	
Toward a Silicon-Based Nuclear-Spin Quantum Computer—Developing the Technology for a Scalable Solid-State Quantum Computer	284
<i>Robert G. Clark, P. Chris Hammel, Andrew Dzurak, Alexander Hamilton, Lloyd Hollenberg, David Jamieson, and Christopher Pakes as told to Jay Schecker</i>	
Fabricating a Qubit Array with a Scanning Tunneling Microscope	302
<i>Marilyn E. Hawley, Geoffrey W. Brown, Michele Y. Simmons, and Robert G. Clark</i>	

John Wheeler

On quantum theory and information

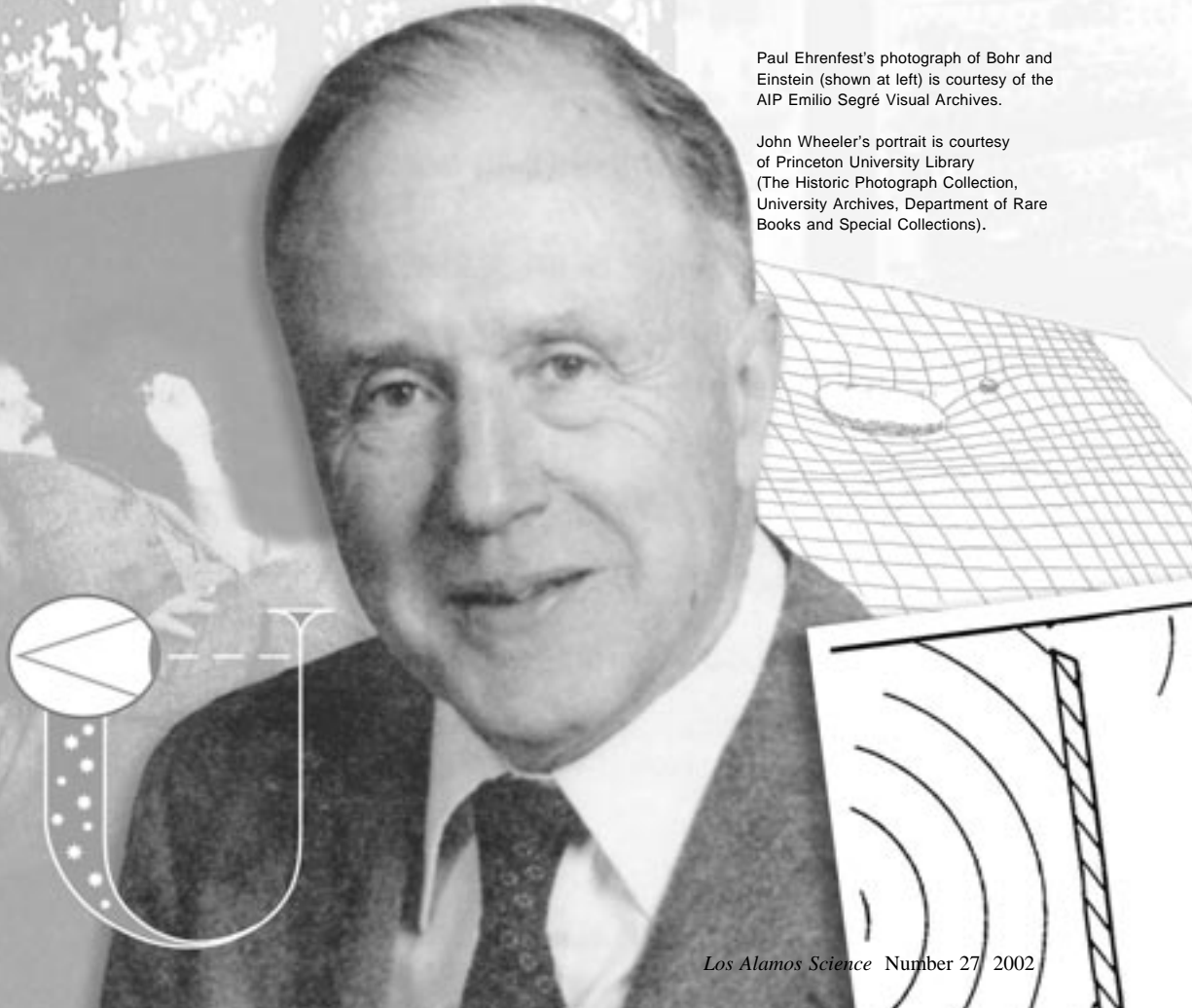
Planck's discovery of the quantum in 1900 drove a crack in the armor that still covers the deep and secret principle of existence. In the exploitation of that opening we are at the beginning, not the end.

—1982

Nothing [in quantum theory]... was more startling than Heisenberg's uncertainty principle, which denied the possibility of simultaneously measuring certain properties of motion. The uncertainty principle introduced us to quantum fluctuations, revealing empty space to be in fact a cauldron of activity.

If the world "out there" is writhing like a barrel of eels, why do we detect a barrel of concrete when we look? To put the question differently, where is the boundary between the random uncertainty of the quantum world, where particles spring into and out of existence, and the orderly certainty of the classical world, where we live, see, and measure? This question...is as deep as any in modern physics. It drove the years-long debate between Bohr and Einstein. . . . Every physical quantity derives its ultimate significance from bits, binary yes-or-no indications, a conclusion which we epitomize in the phrase, it from bit.

—1998



Paul Ehrenfest's photograph of Bohr and Einstein (shown at left) is courtesy of the AIP Emilio Segré Visual Archives.

John Wheeler's portrait is courtesy of Princeton University Library (The Historic Photograph Collection, University Archives, Department of Rare Books and Special Collections).

It is wrong to think of that past [ascribed to a quantum phenomenon] as “already existing” in all detail. The past is theory. The past has no existence except as it is recorded in the present. By deciding what questions our quantum registering equipment shall put in the present we have an undeniable choice in what we have the right to say about the past.
—1980

I have been led to think of analogies between the way a computer works and the way the universe works. The computer is built on yes-no logic. So, perhaps, is the universe. Did an electron pass through slit A or did it not? Did it cause counter B to click or counter C to click? These are the iron posts of observation. Yet one enormous difference separates the computer and the universe—chance. In principle, the output of a computer is precisely determined by the input. Chance plays no role. In the universe, by contrast, chance plays a dominant role. The laws of physics tell us only what may happen. Actual measurement tells us what is happening (or what did happen). Despite this difference, it is not unreasonable to imagine that information sits at the core of physics, just as it sits at the core of a computer.
—1998

Wheeler, J. A. 1980. Beyond the Black Hole. In *Some Strangeness in Proportion*.

Edited by H. Wolf. Reading, MA: Addison-Wesley.

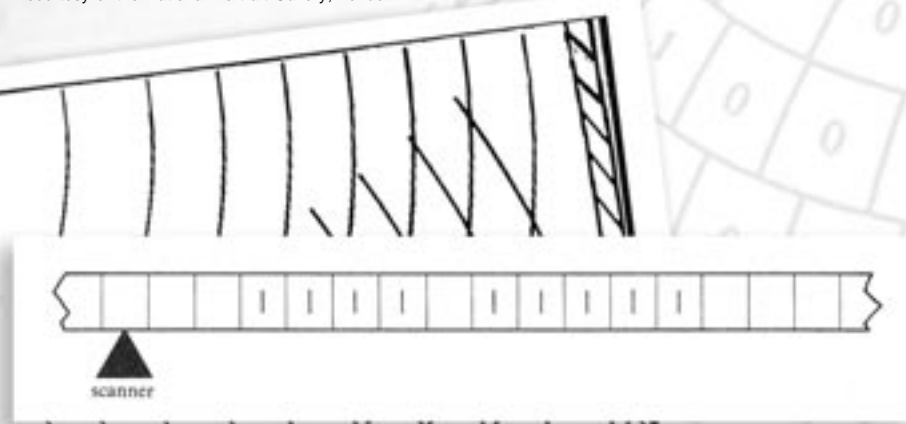
———. 1982. The Computer and the Universe. *Int. J. Theor. Phys.* **21** (6/7).

Wheeler, J. A., and K. Ford. 1998. It from Bit. In *Geons, Black Holes & Quantum Foam*.

New York: W. W. Norton & Company, Inc.

Drawings by John Wheeler surround his portrait. At upper right, matter (the large stone) tells space-time how to curve, and space-time tells matter (the pebble) how to move. The waves from two slits are shown to interfere (below). At lower left is the Eye of the Universe. These drawings and additional images in the background are from *Geons, Black Holes, and Quantum Foam* (see reference above).

The drawing of Alan Turing's automatic adding machine (shown below) is from *Alan Turing, The Enigma* (see reference at lower right), and his photograph (at right) is courtesy of the National Portrait Gallery, London.



Two Giants of Classical Information Theory

If we do not wish to admit that the Second Law has been violated, we must conclude that the intervention which establishes the coupling between [the measuring instrument and the thermodynamic system] must be accompanied by a production of entropy.

—Leo Szilard, 1929, On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings, *Zeit. Phys.* **53**: 840.



Computing is normally done by [a person] writing symbols on paper. . . . I assume that the calculation is carried out on one-dimensional paper, i.e., on a tape divided into squares. I shall also suppose that the number of symbols . . . is finite . . . The behaviour of the computer at any moment is determined by the symbols which he is observing, and his ‘state of mind.’ . . . We may suppose . . . the number of states of mind which need to be taken into account is finite. . . . the use of more complicated states of mind can be avoided by writing more symbols on the tape. . . . Every [simple] operation consists of some change in the physical system consisting of the computer and his tape. [And so, Alan Turing begins to describe his automatic machine that can perform all possible deterministic algorithms.]

—Alan Turing, 1937, On Computable Numbers with an Application to the Entscheidungsproblem, *Proc. Lond. Math. Soc.* **2**: 42. (Excerpts reprinted in Andrew Hodges’ *Alan Turing: The Enigma*, New York: Simon and Schuster, 1983.)



Richard Feynman

On quantum physics and computer simulation

. . . there is plenty of room to make [computers] smaller. . . nothing that I can see in the physical laws . . . says the computer elements cannot be made enormously smaller than they are now. In fact, there may be certain advantages.

—1959

Might I say immediately . . . we always have had a great deal of difficulty in understanding the world view that quantum mechanics represents. . . I cannot define the real problem, therefore I suspect there's not a real problem, but I'm not sure there's no real problem.



I mentioned . . . the possibility . . . of things being affected not just by the past, but also by the future, and therefore that our probabilities are in some sense “illusory.” We only have the information from the past and we try to predict the next step, but in reality it depends upon the near future . . . I'm trying to get . . . you people who think about computer-simulation possibilities to . . . digest . . . the real answers of quantum mechanics and see if you can't invent a different point of view than the physicists . . .

. . . the discovery of computers and the thinking about computers has turned out to be extremely useful in many branches of human reasoning. For instance, we never really understood how lousy our understanding of languages was, the theory of grammar and all that stuff, until we tried to make a computer which would be able to understand language . . . I . . . was hoping that the computer-type

thinking would give us some new ideas . . .

. . . trying to find a computer simulation of physics seems to me to be an excellent program to follow out. . . the real use of it would be with quantum mechanics. . . Nature isn't classical . . . and if you want to make a simulation of Nature, you'd better make it quantum mechanical, and by golly it's a wonderful problem, because it doesn't look so easy.

—1981

Feynman, R. 1959. There's Plenty of Room at the Bottom. Talk given at the annual meeting of the American Physical Society at Caltech. (Excerpt reprinted with permission from Caltech's *Engineering and Science*.)

———. 1981. Simulating Physics with Computers. Keynote address delivered at the MIT Physics of Computation Conference. Published in *Int. J. Theor. Phys.* **21** (6/7), 1982. (Excerpts reprinted with permission from the *International Journal of Theoretical Physics*.)

On “tiny computers obeying quantum mechanical laws”

. . . although I have done mostly physics, from time to time I pay attention to computers. Two years ago Carver Mead . . . discussed with us [that] there ought to be physical laws about the limits in computer design. . . . I got interested in the problem [the amount of heat generated by an operating computer] and worked it all out. It turned out that Charlie Bennett from IBM had worked it all out five years earlier. . . . if you have a reversible machine, the minimum energy requirement is essentially zero. . . . you can have millions and millions of primitive elements doing the calculation, but if the answer has only 40 bits then $40 kT$ is the minimum energy needed.

. . . An exciting discovery, made mostly by Fredkin, was that you can make a computer solely out of reversible primitive elements. . . . With one primitive element [the Fredkin gate] we produce all the effects we need. In addition, the Fredkin gate is reversible. . . . [and therefore] reversible computation is possible. . . .

The next question was what are the limits in computers due to quantum mechanics? . . . What I hoped to do was to design a computer in which I knew how every part worked with everything specified down to the atomic level. In other words I wanted to write down a Hamiltonian for a system that could make a calculation. Then I could calculate the various effects of the limits due to quantum mechanics.

Now, we can, in principle make a computing device in which the numbers are represented by a row of atoms with each atom in either of the two states. That’s our input. The Hamiltonian starts “Hamiltonianizing” the wave function. . . . The ones move around, the zeros move around . . . Finally, along a particular bunch of atoms, ones and zeros . . . occur that represent the answer.

Nothing could be made smaller . . . Nothing could be more elegant. No losses, no uncertainties, no averaging. But can we do it? . . . how can I make the dynamics of quantum mechanics generate a long sequence of unitary matrices? . . . It has been suggested [by Paul Benioff, we believe] that [each unitary] operation . . . can be represented as the action of some Hamiltonian for a definite amount of time. . . . That’s an awful lot of external machinery. . . Let’s get all the atoms into the system. . . [And so, inspired by the ballistic models of Fredkin and Toffoli, Feynman designed a model of a quantum computer in which spin waves would travel through the device to monitor the computational progress. It was the first model after Paul Benioff’s].

—1983

Feynman, R. 1983. Tiny Computers Obeying Quantum Mechanical Laws. Talk delivered at Los Alamos National Laboratory. Published in *New Directions in Physics: The Los Alamos 40th Anniversary Volume*. 1987. Edited by N. Metropolis, D. M. Kerr, and G.-C. Rota. Orlando, FL: Academic Press, Inc. (Excerpts reprinted with permission from the publisher.)



Preface

We live in a quantum world, in which probabilities, not certainties, govern what we see at the submicroscopic level. Interpretations of this fact have been the subject of endless debate since the formulation of quantum theory in the 1920s. One thing, however, is new: In the past decade, we have become increasingly familiar with quantum states. Indeed, at Los Alamos and other laboratories across the globe, individual quanta are being manipulated in ways only dreamt of before.

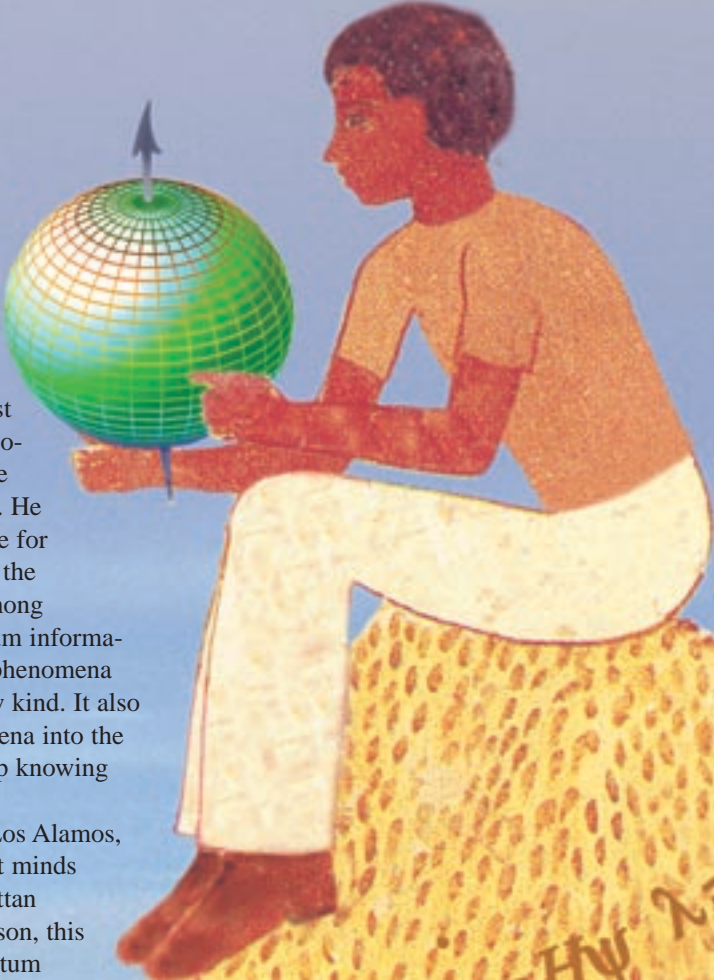
Those efforts have recently intensified as scientists are exploiting a newly identified aspect of the quantum world. It is called quantum information. Its smallest unit is the qubit, a two-level quantum system that can be measured to reveal a “yes” or “no” answer to a question. Thus, measurement of a qubit yields one classical bit of information. Under appropriate conditions, many systems behave as qubits. The polarization states of a single photon, spin-half nuclei in NMR experiments, and a system composed of two relatively stable levels of an ion are among the types of qubits explored at Los Alamos.

Unlike a classical bit, a qubit can be in a pure, yet superposed, state, in which it occupies both levels simultaneously. When measured, a superposition of the two levels behaves like a classical “probabilistic bit,” or pbit, yielding random yes or no answers according to the probability law associated with the particular measurement. The law has the following generic form: p is the probability of measuring yes, and $(1 - p)$, of measuring no. Neither the state of a pbit (that is, its probability law), nor the state of the qubit can be determined from a single measurement. Instead, an infinite sequence of measurements on independent but identically prepared copies of the system is necessary. However, when a qubit is prepared in a pure state, it has a property unknown in the classical world: There will always be one and only one independent property whose measurement will produce a yes answer with certainty, that is, with probability $p = 1$. Moreover, if that property is known, the probability laws associated with all possible measurements on the pure state are also known. In general, the pure states of a quantum system make up a complex projective Hilbert space, which for a qubit, can be pictured as points on the surface of a sphere. That is why qubits are often represented as vectors pointing along directions of a sphere.



In the illustration at right, a young man holds a qubit in his hands. His perspective rests on knowledge accumulated over the last century in quantum physics, information theory, and computer science, the fields that gave birth to the concept of quantum information. He symbolizes the potential of this new resource for communication and computation, as well as the curiosity and excitement it has generated among young men and women. Research on quantum information holds the promise of making quantum phenomena subject to control and manipulation of a new kind. It also holds the promise of bringing these phenomena into the classroom, where young people will grow up knowing the quantum first hand.

Inspiration is derived in many ways. At Los Alamos, a sense of history and the legacy of the great minds who were leading participants in the Manhattan Project are a continuing source. For that reason, this volume about the Los Alamos effort in quantum information and quantum science opens with thought-provoking words from John Wheeler and Richard Feynman (see pages vi–ix). Both were Manhattan Project pioneers, and as discussed below, both have helped launch the field of quantum information science and renew interest in the foundations of quantum theory and measurement.



$$E = \hbar\omega \quad i\hbar\psi = H\psi \quad \lambda = h/p$$

$$|\langle ab \rangle - \langle ac \rangle| \leq 1 - \langle bc \rangle \quad S = -\text{Tr} \rho \ln \rho$$

$$\Delta x \Delta p \geq \hbar/2 \quad (\alpha|\uparrow\rangle + \beta|\downarrow\rangle)|0\rangle = \alpha|\uparrow\rangle|0\rangle + \beta|\downarrow\rangle|0\rangle$$

$$\text{cnot} = |0\rangle_A A|0\rangle + |1\rangle_A A|1\rangle (|0\rangle_B B|1\rangle + |1\rangle_B B|0\rangle)$$

$$p = \sum_i p_i |w_i\rangle \langle w_i| \quad E = \hbar\omega$$

The Strangeness of the Quantum World

“The elementary quantum phenomenon is the strangest thing in this strange world. It is strange because it has no localization in space and time. It is strange because it has a pure yes-no character—one bit of meaning. It is strange because it is more deeply dyed with an information-theoretic flavor than anything in all physics.”

—John Archibald Wheeler (1984)

Wheeler is best known for working out the theory of nuclear fission with Niels Bohr in 1939 and for pioneering black-hole physics in the 1950s and 1960s. He has also spent well over half a century inspiring his many students and associates to think “outside the box.” Together with Feynman, his graduate student in the early 1940s, Wheeler explored his “crazy” idea of treating particle trajectories going forward and backward in time on an equal footing. Both that experience and Dirac’s ideas influenced the calculational shorthand known as Feynman diagrams and Feynman’s formulation of quantum electrodynamics, for which Feynman received the Nobel Prize. In the 1960s and 1970s, Wheeler continually probed the connection between physics and information and opened the way for his graduate students and younger colleagues to help create a new field.

Quantum theory teaches us that, on the smallest scales, nature is observed to be granular. Electromagnetic radiation is absorbed and radiated in discrete units, which we call photons. The stable energy levels of an atom are also discrete, and electrons can be seen to go from one level to the next by “quantum jumps.” The counterpoint to this ubiquitous discreteness is a form of continuity even more challenging to our everyday experience: Individual quantum systems can exist

in a superposition of different states, corresponding, for example, to photons traveling along different paths in the famous double-slit experiment. Through measurement, the photon state, or wave function, “collapses” and becomes concentrated at the spot where it is observed, but repeating the measurement on another identically prepared photon typically produces a different, though equally definite, outcome. The state of each identically prepared photon is what determines the probability of obtaining different outcomes in many such repetitions and for many such measurements. To physicists imbued with the realistic local worldview of classical physics,

the result is indeed surprising. Like a wide wavefront that has found its way through both slits simultaneously, each photon interferes with itself. Yet, when measured, as if by magic, each reduces to a point of light at some random location on the screen. The familiar interference pattern, predicted by both classical electromagnetic theory and quantum theory, arises when many photons are looked at together or in sequence (see the box “The Double-Slit Experiment” on page 142). The single photon—spread throughout space and observed only at a point in space—challenges our very concept of position as an attribute of the particle.

Viewed differently, in the delayed-

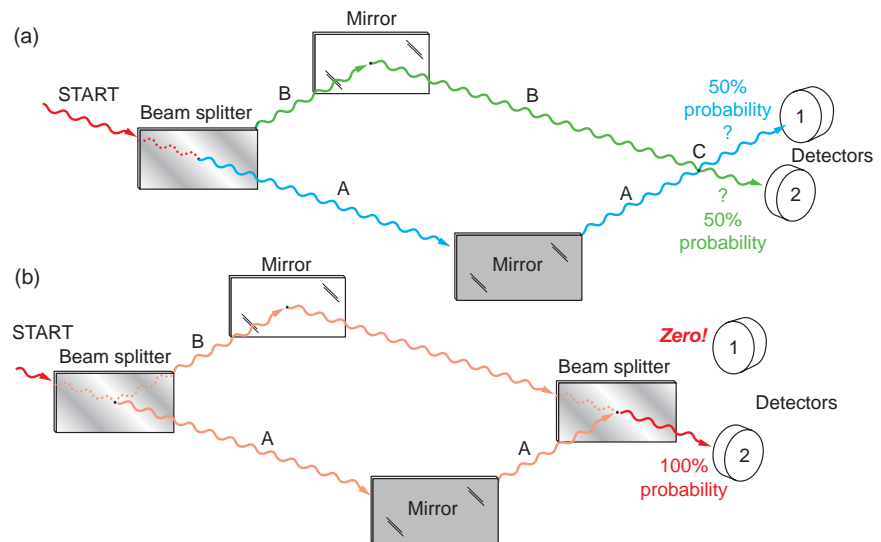


Figure 1. The Delayed-Choice Experiment

(a) At START, an incoming light wave encounters a beam splitter, which splits it into two beams of equal intensity. Each is reflected by a mirror and the two cross paths at point C. Detectors located past point C tell by which route an arriving photon has come. (b) The arrangement is the same as in (a) except that now a beam splitter is inserted at point C. It brings beams A and B into destructive interference on one side, so that detector 1 never registers anything, and into constructive interference on the other, so that every photon entering at START is registered at detector 2 in the idealized case of perfect mirrors and 100 percent photodetector efficiency. In (a), one finds out by which route the photon came. In (b), one has evidence that the arriving photon came by both routes. In the “delayed-choice” version of the experiment, one decides at the very last picosecond whether to insert the second beam splitter. In other words, one waits until the photon has done most of its travel before deciding whether the photon “shall have come by one route or by both routes.” (Diagram adapted with permission from John Wheeler, “Law without Law,” in *Quantum Theory and Measurement*, edited by J. A. Wheeler and W. H. Zurek, Princeton, NJ: Princeton University Press, 1983.)

choice experiment (Figure 1), the photon's behavior challenges our naive concept of causality. In (a), a photon hitting a beam splitter will follow path A or B, arriving at detector 1 or 2, respectively, with equal probability. One can deduce that, in this arrangement, the photon has followed a definite path: If either path is blocked, the count in the corresponding detector drops to zero. In (b), the setup is the same as in (a) except that a beam splitter is inserted at C, the point where the two paths cross. Now, interference causes all photons to arrive at detector 2 and none at detector 1. The photon's ability to traverse both paths is alone responsible for this situation: With either path blocked, the photons reach each detector equally.

We can turn (a) into (b) by inserting a beam splitter at C, and we can choose whether to insert it at the very last moment. In this way, we can control whether the photon behaves as if it had taken one path or the other or had traveled along both paths. Now comes the contradiction to a local realist's view of causality: The beam splitter can be inserted after the photon is done traversing the region in question!

These paradoxes led Wheeler to view our physical reality through the lens of information theory: "Every item of the physical world has at bottom an immaterial source . . . what we call reality arises in the last analysis from posing yes-no questions and the registering of equipment-evoked responses; . . . in short, all things physical are information-theoretic in origin."

The link between what quantum mechanics tells us might happen—"multiple paths, interference patterns, spreading clouds of probability"—and what does indeed happen in the observable world is provided by the measurement process and/or the participation of the observer. The late Rolf Landauer of IBM, sometimes called the conscience of the physics of

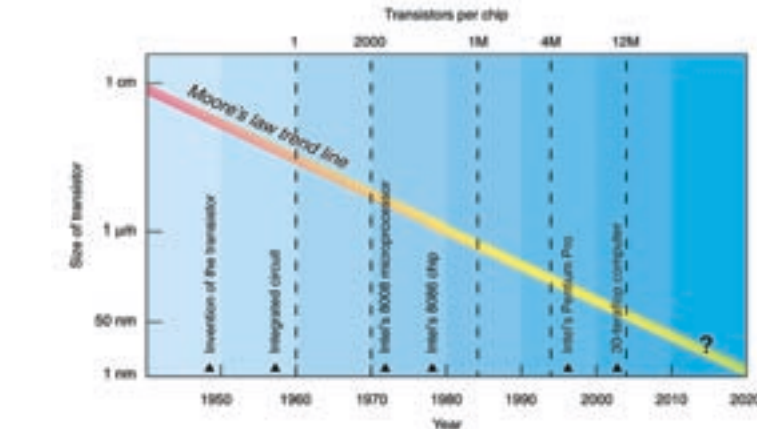


Figure 2. Miniaturization of the Transistor

By 1980, over 100,000 transistors were on a single chip. Today, that number is 40 million, or 20 million per centimeter squared. Quantum effects will become important when the size of the transistor approaches the nanometer scale and only a few electrons are involved in determining current flow.

information, echoed Wheeler's view (1999): "I am suggesting that, contrary to our prevailing views, the laws of physics did not precede the universe and control it, but are part of it. Wheeler has stated that the laws of physics result from quantum measurement on the universe." These mind-bending thoughts about the elementary quantum phenomenon and the fundamental role of measurement, or information processing, in determining the laws of physics can be turned around to ask another profound question, "how does the nature of physical law limit our ability to process information?" Both questions were the subject of a seminal meeting in 1981.

Physical Limits on Computation and the First Models of a Quantum Computer

About 60 physicists and computer scientists gathered at the workshop "Physics of Computation" sponsored by the Massachusetts Institute of Technology (MIT). The organizers were Rolf Landauer, Tom Toffoli, and Ed Fredkin. The participants included

Wheeler, Feynman, Charles Bennett, and Paul Benioff. According to the organizers, they all shared the belief that "physics and computation are interdependent at a fundamental level."

A general concern at that time was the drive toward ever-increasing computer power through miniaturization of components. Moore's law—the doubling of transistor density on a chip every eighteen months—had been describing an ongoing trend for several decades (see Figure 2). Because transistor elements were getting smaller, more and more of them were being crammed onto a chip, proportionately increasing computing power.

In the foreseeable future, each element would shrink to a size at which quantum effects become important. The question is how small could each get? Would the heat generated from so many computational steps in a tiny area lead to a literal meltdown of the chips? Could one use quantum mechanical elements to build computers—single atoms, perhaps? Would the time-energy uncertainty relation dictate the rate of energy dissipation? Would quantum fluctuations get in the way of reliability?

The research staff at IBM had

thought about these questions for many years and made some important strides. Landauer, for example, had repeatedly emphasized that information is always physical. He had delved deeply into the physics of information processing and in 1961 understood that erasure, the discarding of information, is an irreversible process that produces heat and increases entropy. He also assumed that computation necessarily involves erasure.

In 1973, Bennett showed that assumption to be false. Building on Szilard's work that connected information and entropy and Landauer's insight that erasure is the problem, he developed a logically reversible model of a Turing machine. This formal machine model of a universal computer had a memory tape, read-write head, and finite-state internal machine in the style of a Turing machine (see the box "The Universal Turing Machine" on page xvi). Bennett managed to design a reversible one-to-one mapping of information from input to output by employing three tapes instead of one: The first tape was for the input data; the second, for a history of intermediate results; and the third, for keeping a copy of the output. Because all operations would be done reversibly, the machine could run backwards, thereby retracing its steps, disposing of the intermediate results along the way, and returning to its initial state. This logical reversibility implied that, in principle, one could construct a thermodynamically reversible physical machine, which if run slowly, could perform any computation reversibly with arbitrarily little energy dissipation per step. Thereby, Bennett had found a way around the heat problem, but at the expense of speed. What, if any, were the limits quantum mechanics would place on computation?

It was Benioff who first showed that reversible computation with no

dissipation could be realized very naturally in a computer made of quantum mechanical parts. In 1980, he had begun developing quantum mechanical models of computation as a first step toward a model of intelligent systems. This very first model of a quantum computer consisted of a lattice of spin-half atoms that would evolve smoothly and deterministically according to the Schrödinger equation of quantum mechanics. Benioff invented an appropriate spin Hamiltonian that would govern the dynamics of this spin system, he proposed that the Hamiltonian act for a specific period to accomplish speci-



"Information is inevitably tied to a physical representation and, therefore, to all the possibilities and restrictions allowed by our real physical universe. . . . This is the viewpoint invoked by [Leo] Szilard. . . . His understanding of the physical nature of information was truly pioneering."

—Rolf Landauer (1999)

fied operations, and he showed that the states of the system would evolve with time, as needed to carry out the basic logic operations of a Turing machine. Because quantum mechanical time evolution is unitary, it generates a one-to-one reversible mapping of the system from one state to the next that can implement computational steps with no dissipation. Benioff presented his model at the 1981 "Physics of Computation" workshop. His ideas were revolutionary at the time. Many scientists had believed that any fast

switching event would, by the time-energy form of the Heisenberg uncertainty principle, require a minimal energy expenditure, and therefore they expected to find intrinsic limits to the speed and accuracy in a computer obeying the laws of quantum mechanics.

Benioff showed that this fear was unfounded: The laws of physics place no upper bound on the speed attainable or lower bound on energy dissipation during computation. The true significance of the uncertainty principle is, however, that the speed would be limited by the particular quantum dynamics of the computer: That is, the time per operation is limited by the Hamiltonian (energy) in the system divided by \hbar . Furthermore, the size of the elements could be reduced to individual atoms, or as we will see below, individual photons. Of course, although Hamiltonian evolution was simple to describe theoretically, Benioff had not dealt with the practical issues such as creating the initial state of the system, reading out the answer, the probabilistic nature of the quantum mechanical answer, and keeping the system isolated from the environment.

In "Zig-Zag Path to Understanding," Landauer recalls Benioff's 1981 presentation and the reaction to it: "[My own] attempts to produce a quantum version of the reversible Bennett-Fredkin-Turing machine had gotten hopelessly bogged down . . . Benioff saw the way to do that. You invoke a Hamiltonian (or a unitary time evolution) that causes the information-bearing degrees of freedom to interact, and to evolve with time, as they do in a computer. You introduce no other parts or degrees of freedom. . . . Feynman was present at the 1981 workshop at MIT, where many of us discussed Benioff's notions. . . . Did we understand Benioff? Feynman did not need much of a clue, and as a result generated his own very appealing and

Quantum Issues at the 1981 Workshop as Remembered by Paul Benioff

“During the 1960s and 1970s, there was much interest in making fast, more powerful computers by miniaturizing components and packing more computer power into smaller volumes of space and time. However, there were two main problems: One was the appearance of quantum mechanical effects and the other was the generation of heat due to the irreversibility of the computation process. Until the work of Bennett in 1973, it was thought that the computation process was necessarily irreversible, with energy dissipation associated with information erasure. However Bennett showed that to every irreversible computation there exists an equivalent reversible computation.”



“Yet Bennett’s work did not address concerns related to quantum effects. Here, the concerns were twofold. One was that the energy-time uncertainty principle meant that the amount of energy dissipated per computation step was bounded below by Planck’s constant divided by the switching time. However, as Landauer

pointed out in 1982, the uncertainty principle does not mean that the energy is necessarily dissipated. I used this fact implicitly in my models. They operated at the quantum limit in that the total energy of the system was given by the energy-time uncertainty principle, but that energy was not dissipated.”

“The other concern was that computation steps of a conditional nature—if a system is in state 0, do this; if it is in state 1, do that—necessarily involved measurement, which of course, does dissipate energy. The view that reading is equivalent to measurement is again erroneous. It ignores the fact that measurement consists of two stages: first establishing a correlation between states of the measured system and the apparatus, that is, an entangled state of the system and the measuring apparatus, and second, amplification or decoherence. It is this latter stage of decoherence, much studied and developed by Wojciech Zurek, that leads to dissipation. However, only the first step is necessary in quantum mechanical models of computation. This step, which does not dissipate energy, was used implicitly in my models and is an essential part of quantum computation models used today.” (Private communication)

effective view of quantum mechanical computation” (Landauer 1994).

The keynote address at the 1981 workshop was delivered by Feynman, and it too was to have a profound impact on the community. The topic, tangential to the rest, was the problem of simulating physics with computers—in particular, simulating quantum physics. Feynman told his audience that this topic had a twofold interest: “learning something about the possibilities of computers, and also something about possibilities in physics.” This interest was fueled by his close association with Fredkin, a proponent of the idea that space and time are discrete, not continuous, and that the Universe is, in essence, a giant digital computer.

Feynman analyzed the problem with his typical flare and brilliance. He limited the computer to one with

local interconnections and the type of simulation to one in which the number of computer elements required to simulate a large physical system is proportional to the space-time volume of the physical system. “. . . [C]lassical physics is local, causal, and reversible, and therefore apparently quite adaptable to computer simulation,” provided, Feynman said, that we allow space-time to be discrete. In quantum mechanics, however, “we know immediately that we get only the ability, apparently, to predict probabilities . . .”

Could a system of probabilistic universal computers, classical Turing machines supplemented with random number generators, simulate the probabilistic world of quantum mechanics? His answer was a resounding “NO!” A probabilistic computer could not reproduce events with the same proba-

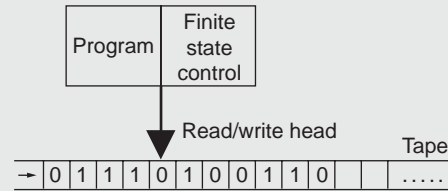
bilities observed for quantum mechanical systems, without, in essence, simulating the entire universe at each point. A computer with only local interactions and polynomial resources would have to solve the famous hidden variable problem to match quantum probabilities, but John Bell had shown that only a nonlocal theory could predict the same probabilities as quantum mechanics (see the box “The EPR Paradox and Bell’s Inequalities” on page xviii).

Feynman also concluded that such a classical computer could not simulate the wave function of a quantum system of N particles because the number of variables needed to describe the problem would grow faster than exponentially with N . He suggested, however, that “You can simulate it with quantum computer elements. It’s not a Turing machine but

The Universal Turing Machine

In 1935, Alan Turing imagined a machine that would be capable of answering any question that could be answered logically. His invention was a blueprint for the modern programmable computer, and he proved theoretically that it could perform any computation that could be carried out through logical manipulations. The Turing machine has three elements (see figure): (1) an internal machine *L* that contains the program and can assume any one of a finite number of states, (2) a computation tape containing an infinite number of cells that serves as the memory, and (3) a read-write head that scans the tape, one cell at a time performs read/write operations on the cells, and can shift one cell to the left or right, or stay in place, depending on the contents of the cell, the state of the internal machine *L*, and the program instruction. The read/write alphabet is finite, say, zero and one, and it also includes a blank and a start symbol. Although operations such as addition require many steps, the machine is very powerful. The Church-Turing thesis states the following: The class of functions computable by a Turing machine corresponds

exactly to the class of functions that we would naturally regard as being computable by any algorithm (definite procedure). Turing’s invention was built on



the insight of Kurt Gödel that both numbers and operations on numbers can be treated as symbols in a syntactic sense. Today, we take for granted that all information, including programmable instructions, can be expressed by strings of ones and zeros (or “yes” and “no” answers) and that all computations, from simple arithmetic to proving of abstract theorems, can be accomplished when a small set of mechanical operations (the program) are applied to these bit strings in some specified order.

a machine of a different kind.” Feynman then guessed that “every finite quantum mechanical system can be described exactly, imitated exactly, by supposing that we have another system such that at each point in space-time this system has only two possible base states. Either that point is occupied or unoccupied—those are the two states.” In other words, a universal quantum simulator, closely resembling today’s universal quantum computer, could be used to simulate discrete quantum systems. That idea is being pursued today. Only later, after Benioff’s presentation at the 1981 workshop, did Feynman develop his own model of a universal quantum computer. It included a system for monitoring within the computer the progress of the calculation so that one would know the endpoint of the calculation and the time at which to read

out the answer. At all decision points, however, this computer was in a definite state; never was superposition of different computational histories used as a tool.

In the spring of 1983, on the 40th anniversary of the Los Alamos National Laboratory, Feynman returned to Los Alamos for the first time since the 1940s. He joined his colleagues from the Manhattan Project era in a seminar on forward directions in physics. Feynman talked about reversible computing and his own model for a quantum computer in a talk entitled “Tiny Computers Obeying Quantum Mechanical Laws.” Feynman’s model was not the first, and it is not the model used in today’s theoretical and experimental studies. Nevertheless, it stands as a record of Feynman’s immense interest in this emerging area.

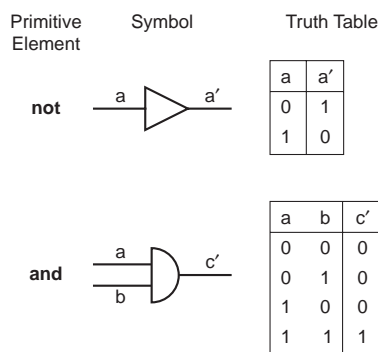
Physical Realizations of Classical vs Quantum Computers

One of the marvels of modern life is that the “universal” computer—any machine that is as powerful as a Turing machine—has become totally commonplace. All desktop computers are universal in the sense that, given enough time and memory, they can do any computation that is done on any other computer no matter how large. The model used to construct modern computers and the one used in this volume to describe quantum information processing are circuit models. For a classical computer, this model is realized as a set of physically distinct logic gates, usually implemented as transistor elements, connected by real wires. Information entered in the input register is processed as electric

currents that go through a prescribed set of logic gates whose voltages are set according to the program instructions. Results are recorded in an output register. The figure at right shows two of the standard logic gates: the **not** gate and the **and** gate. Both the electronic symbols and corresponding truth tables for those operations in binary notation—0 = false (no), and 1 = true (yes)—are shown. These are the only gates needed to construct a universal computer that can perform all possible computations: In fact, a **nand** gate, constructed as an **and** gate followed by a **not** gate, suffices. Note, however, that the **and** (and **nand**) gate is obviously irreversible—one cannot determine the identity of the two inputs from the single output. A fully capable computer also needs fanout, the ability to send the same output to multiple inputs, and it needs to perform iteration (known as loops or recursion).

In the physical realization of reversible computing achieved with a

quantum computer, on the other hand, there are no real wires. The input and output register is the same set of qubits, a row of, say, spin-half atoms in an ion trap, in a molecule, or embedded in a solid matrix. The “wires” car-



rying the qubits from one gate to the next are their time lines, and the logic gates are a sequence of unitary operators (typically external radio-frequency pulses and evolutions due to the internal interaction Hamiltonian of the system) that change the states of the qubits (see Figure 3). During the computation, the quantum mechanical

wave function for the system evolves smoothly and deterministically according to the Schrödinger equation.

Once the computation is complete, the answer is obtained by a measurement, and, hence, is often probabilistic. A reliable answer typically requires repeated computations. It is, however, possible to design efficient quantum algorithms so that the final answer in the qubits is close to deterministic—any one measurement has sufficient information to allow extracting the desired answer with high probability. This deterministic feature is illustrated for the parity problem (introduced on page 21 of the primer) and Shor’s algorithm (see the article “From Factoring to Phase Estimation” on page 38). The length of time for a quantum computation is limited by the intrinsic relaxation time of the system (various internal interactions can drive the two-level qubits to the ground state) and the decoherence time—the gradual “leakage” of quantum coherence to the environment.

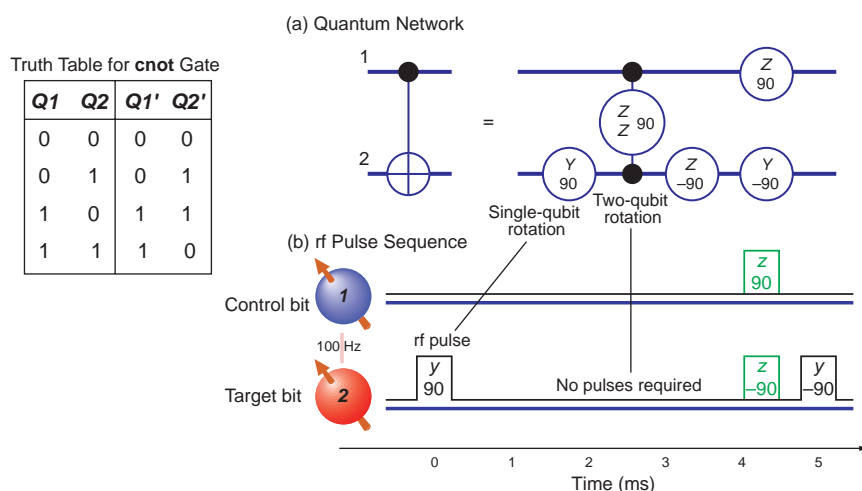


Figure 3. The cnot Gate in an NMR Quantum Processor

In quantum computers, the cnot (controlled not) gate plays a role analogous to that of the nand gate in classical computers. Part (a) shows the NMR quantum circuit, the sequence from left to right of 1-qubit rotations and 2-qubit internal Hamiltonian evolution, that executes the logic of the cnot gate, namely, reverse the spin of qubit 2 only if qubit 1 is in the 1 state (see Truth Table). Part (b) shows the rf pulse sequence needed to execute the cnot gate. Note that the two-qubit operation occurs by allowing the internal spin-spin Hamiltonian to evolve the system for a specified period. (See the article “NMR and Quantum Information Processing” on page 227.)

The Unique Properties of Quantum Information

Feynman’s notion that any finite quantum system could be simulated by a device made of spin-half atoms expanded the scope of what one might do with a computer made of quantum mechanical elements. In 1985, Deutsch took this idea one step further, suggesting that a computer made of elements obeying quantum mechanical laws could efficiently perform certain problem-solving and computational tasks for which no efficient classical solution was known. The key features of quantum mechanics to be exploited were the principle of linearity, which allows the components of a superposition of multiparticle states to evolve simultaneously, and the principle of interference, which allows certain superpositions

Continued on page xx

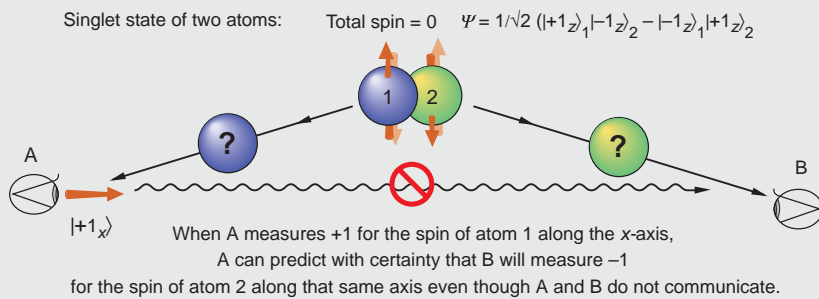
The EPR Paradox and Bell's Inequalities

When quantum theory was first formulated, Albert Einstein, Nathan Rosen, and Boris Podolsky questioned its completeness. They described a situation in which a predictable outcome could not be predicted by quantum mechanics. David Bohm illustrated this paradox (called EPR) using a molecule of two spin-half atoms with total angular momentum zero, that is, a coherent superposition of two product states (atom 1 in spin up) \times (atom 2 in spin down) and (atom 2 in spin up) \times (atom 1 in spin down)—see Figure A. For this maximally entangled state (defined in the

the system, then the measurement result defines a 'real' property. Because the spin direction of atom 2 can be predicted with certainty, spin direction must be a 'real' property of the atom. Hence, if quantum theory were complete, it would predict the spin direction of each atom independently. In other words, "since the initial quantum mechanical wave function does not determine the result of an individual measurement, this predetermination (of the spin direction of atom 2) implies the possibility of a more complete specification of the state" (Bell 1964). To a local realist,

imagine that the initial quantum state can be mimicked classically by two arrows pointing to random, but opposite, points on a sphere. When the system splits apart, each arrow has an equal probability of going to the right or left. With the simple local rules given in parts (i) and (ii) of Figure B, a classical theory can predict the perfect correlations seen when both spins are measured along the same axis—the spins point in opposite directions with probability 1. Only when one measures the two arrows (spins) in different directions, say, z and d , does the local realistic theory contradict the quantum results—see Figure B(iii). Clearly, classical analogues of entangled qubits need to be more complicated than a pair of arrows; as long as the results of measurements along all axes depend on a single variable, such as the arrow's direction in our example, the quantum mechanical results cannot be reproduced. Basically, qubits would need to be modeled by machines that calculate the results of measurements along different axes using different combinations of hidden variables, all of which may be random but must be correlated between the two measurement paths. This possibility was ruled out by Bell, who showed that the entire set of correlations implied in Figure B cannot be reproduced by any local realistic theory.

Figure A. Is Quantum Mechanics Complete?



main text), quantum mechanics predicts only that, if one atom's spin is measured along an axis chosen arbitrarily, the other atom's spin will always turn out to be its opposite when measured along the same axis. Quantum mechanics also requires, however, that each individual measurement have a random result. One concludes that, if the atoms are split apart and the spin of atom 1 is measured after the two are separated by a large distance, a measurement of the spin of the second atom along that same axis would be completely determined without any signalling from atom 1.

In the worldview of a local realist, a complete theory is one in which every 'real' property of a system can be predicted. Further, if the outcome of a measurement can be predicted with certainty without interfering with

therefore, the property of nonlocal correlation seen in David Bohm's example required introduction of a more complete theory, possibly involving "hidden variables" (or degrees of freedom over which one would have no control) that would determine the outcomes of individual measurements. This apparent incompleteness of quantum theory was one issue in the famous debate between Bohr and Einstein about the validity of quantum mechanics.

Any hope of a more complete theory was laid to rest when John Bell (1964) showed that no local realistic theory could possibly reproduce the probabilities computed according to quantum mechanics, without at some point invoking nonlocal effects. Figure B illustrates the basis of Bell's proof. In that figure, we construct a local realistic theory that matches the results depicted in Figure A. We

In every such classical system, one can ask for the probability $p(z+, d+, \underline{d}+)$ that the hidden variables of the first particle have such values that measuring its spin along the three fixed axes z , d , and \underline{d} (see next page) would yield positive values. Because measurements along the \underline{d} axis can provide only two possible values, positive or negative—one concludes that $p(z+, d+) =$

$p(z+, d+, \underline{d}+) + p(z+, d+, \underline{d}-)$.
On the other hand, one obviously has $p(z+, d+, \underline{d}+) \leq p(d+, \underline{d}+)$ and $p(z+, d+, \underline{d}-) \leq p(z+, \underline{d}-)$.

Putting these together, one obtains the master result that the distribution of hidden variables must satisfy

$$p(z+, d+) \leq p(z+, d-) + p(d+, d+) .$$

Let us now consider an event in which the first spin is measured to be positive along the z -axis, and the second is observed to be negative along the d -axis. Because of the strict antiparallelism of the two spins whenever both are measured along d , we can conclude that, if the first spin had been measured along the d -axis, the measurement would have yielded a positive result. Thus, such events for spins 1 and 2 occur when, and only when, the hidden variable of the first spin is in such a state that it would provide positive results to measurements on both the z - and d -axis. In our notation, such a state for spin 1 happens with probability $p(z+, d+)$.

Thus, the probability of measuring a positive first spin along z and a negative second spin along d , $P_{zd}(+, -)$, is equal to $p(z+, d+)$. Using similar logic, we can transform our master inequality above to a statement about correlations:

$$P_{zd}(+, -) \leq P_{zd}(+, +) + P_{dd}(+, -) ,$$

where $P_{zd}(+, -)$ represents the probability that, in an experiment in which the first spin was measured along z and the second along d , the observed outcomes were positive and negative respectively. This is a particular case of Bell's inequality, which every classical theory model must satisfy.

On the other hand, consider measuring the entangled system of two spin-half atoms along the same axes z , d , and \underline{d} . One can easily obtain the probabilities from quantum mechanics:

$P_{zd}(+, -) = 3/8$, $P_{zd}(+, +) = 1/8$, and $P_{dd}(+, -) = 1/8$, and the inequality is clearly violated by the quantum system. Our classical reasoning led us astray: An entangled state is an indivisible unit, and trying to describe it probabilistically out of local properties assigned to its subsystems, even if they are correlated, is forever doomed to failure.

Bell's result changed forever our understanding of quantum mechanics and led to the modern view of quantum measurement.

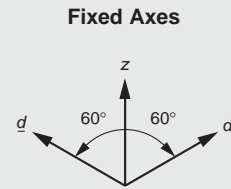
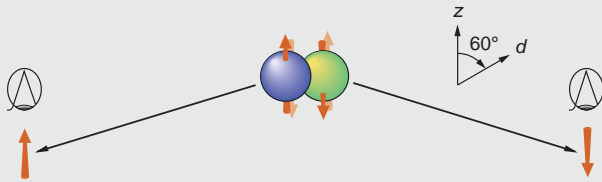


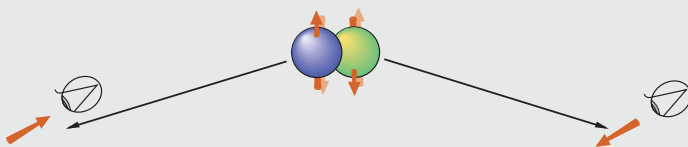
Figure B. Can a Local Realistic Theory Predict Quantum Mechanical Probabilities?

The following are assumptions for this thought experiment: (1) Spins in the initial state are assumed to point along opposite directions. (2) The spins fly off in opposite directions. (3) Each spin is equally likely to go to the left or right.

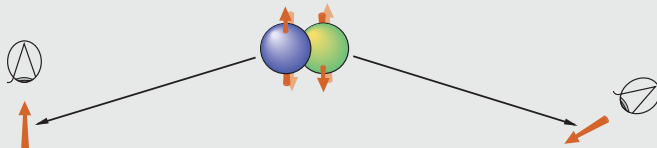
i. Measure both spins along the z -axis. To match quantum correlations, assume that the probability to measure a positive spin $p(z+)$ or a negative spin $p(z-)$ is $1/2$. Both quantum and classical systems yield perfect correlations. The probability of spins pointing in opposite directions is 1.



ii. Measure both spins along the d -axis. To match quantum correlations, add the deterministic rule that + along z is measured as + along d and that - along z is measured as - along d . Again, quantum and classical theory predict spins point in opposite directions with probability 1.



iii. Measure the spin traveling to the left along the z -axis and the one traveling to the right along the d -axis. Quantum and classical correlations do not match!



Measurement Results

Correlation	Quantum	Classical
$P_{zz}(+,+)$	0	0
$P_{zz}(+,-)$	1/2	1/2
$P_{zz}(-,+)$	1/2	1/2
$P_{zz}(-,-)$	0	0

Correlation	Quantum	Classical
$P_{dd}(+,+)$	0	0
$P_{dd}(+,-)$	1/2	1/2
$P_{dd}(-,+)$	1/2	1/2
$P_{dd}(-,-)$	0	0

Correlation	Quantum	Classical
$P_{zd}(+,+)$	1/8	0
$P_{zd}(+,-)$	3/8	1/2
$P_{zd}(-,+)$	3/8	1/2
$P_{zd}(-,-)$	1/8	0

Continued from page xvii

to be detected with high efficiency.

To illustrate these ideas, first consider the properties of a single spin-half system, say, a single particle carrying the smallest nonzero amount of spin allowed by quantum mechanics. The particle behaves as a qubit: Measuring its spin (the amount of its intrinsic angular momentum) along any axis yields one of two quantized values, $+\hbar/2$ or $-\hbar/2$, and as mentioned early on, repeated measurements on independent, identically prepared systems yield a probability law for the two results (that is, the state of a pbit). The system, can, however, be prepared in a pure state with the spin pointing along a definite direction, so that measurement of the spin component along that direction results in $+\hbar/2$ with probability 1.

Suppose also that the system then evolves in isolation. By that we mean that laser pulses and other external sources can cause the spin to change direction in accordance with the laws of quantum mechanics, but because those sources have very large quantum uncertainties, the interaction with the qubit causes almost no change in their states. To put it differently, evolving in isolation means that the qubit can change state through interaction with the external world, but the external world has no information about the spin state of the qubit. Under these conditions, the pure state stays pure: Its spin always points in some definite direction (which, of course, changes with time), and if one happens to measure the spin along that direction or its opposite, one would be guaranteed to obtain a definite result and not disturb the state. Moreover, if one knew the preparation procedure and evolution that led to the state at the time of measurement, one could predict, in one stroke, both its direction and the probabilities for the measurement results along any direction.

For a register of N such qubits, the number of orthogonal pure states, or

states in a complete basis, is exponentially large, 2^N to be precise. Just like a single qubit, however, this N -qubit quantum system can be in any superposition of these exponentially large number of basis states. Furthermore, the linearity of quantum mechanics implies that a sequence of few-qubit unitary operations designed to perform a given computation will do so on any superposition as easily as on a particular one. In this way, it effectively performs an exponentially large number of calculations simultaneously, without needing exponential resources at any stage.

When Deutsch conjectured (1985) that these simultaneous calculations could be exploited to solve problems more efficiently than could be done on a classical computer, he was quick to point out that this “quantum parallelism” is not an analogue of classical parallel computations. In fact, any such computation followed by measurement can yield only one N -bit answer. A direct measurement similar to that in case (a) of Figure 1 would collapse the final state to the results of a single randomly chosen calculation, a calculation that could have been performed on a classical computer equally easily. In contrast, quantum algorithms are carefully designed to be like case (b) of Figure 1; that is, interferences between the results of a large number of simultaneous evaluations are arranged so as to produce definite outcomes. Those outcomes provide information about global patterns (such as the periodicity of a function).

At first glance, this extra ability of quantum computers seems surprising. After all, if the initial superposition of the 2^N basis states is a collection of N pure qubits, each of which can be represented as an arrow pointing in some direction, and the computational steps maintain the purity of these individual qubits, then those steps could be viewed as rules for turning the arrows around. Such rules, it is easily shown,

can be implemented on a classical computer with no difference in efficiency or precision. And if this were all there was to quantum computers, they could be no more powerful than classical ones.

Here, however, is the interesting part: Although quantum computations require only two-qubit operations at each step, many steps together are effectively multiqubit operations. Hence, the individual qubits do not evolve in isolation. Under these conditions, quantum mechanics assures us that only the entire N -qubit register is in a pure state, not the individual qubits; and this is where the miraculous nature of quantum correlations comes in. Many of the pure states of this N -qubit system display a peculiar phenomenon called entanglement: Even though the state of the register is pure—that is, we know as much as the uncertainty principle allows us to know about the system—and the entire system can be conveniently represented as a classical arrow, the states of the constituent qubits are not pure. And so, the state of the whole system is not describable by specifying the state of each qubit separately. An entangled state of more than one qubit is one that cannot be described as a probabilistic mixture of the product of single qubit states; a two-qubit state is called maximally entangled when it is pure, yet provides no information about local measurements on individual constituent qubits. An example of a maximally entangled state is provided in the box “The EPR Paradox and Bell’s Inequalities” on page xviii. Entangled states are more akin to a classical register of probabilistic subsystems in which the interesting information (that is, the results of the calculation so far) is encoded in the numerous correlations between the subsystems. An analogous classical system, without the benefit of the multiparticle superpositions, would have to separately keep track of these correlations,

which build up exponentially fast as the calculation proceeds. Whereas a quantum operation that changes the states of only a few qubits automatically updates the entire multiparticle superposition, the corresponding computational step in the classical system would require updating all these correlations and would become exponentially expensive. Note that it is not the entangled states per se that make quantum computation more efficient than classical analogues. Instead, enhanced computational power is a common feature of general quantum evolutions. Only a computation involving a very limited set of operations has the possibility to be mimicked classically. Conversely, unentangled evolutions of pure states can be mimicked classically because they, of necessity, involve very few kinds of operations. It is an open question whether the larger, but still limited, space of quantum evolutions that do not entangle mixed states of large number of qubits can be simulated efficiently classically, or whether they are powerful enough to perform scalable, useful computations.

Two specific features are responsible for the power of quantum computation: Because quantum mechanics causes multiparticle superpositions to evolve linearly, each computational step can carry out operations that would need an exponential number of classical resources. At the same time, the interference principle allows readout of certain global properties of the results. Those properties are often algorithmically unobtainable without evaluating the computation on each of the exponentially large number of input states. Deutsch's original quantum algorithm gave a solution for one such global property.

The area in which quantum entanglement does serve as a key resource is communication. The idea of exploiting the properties of quantum states for communication was born in the late sixties, when Stephen Weisner invented

a quantum scheme for preventing counterfeiting of paper currency. His scheme was based on two properties of single quanta in pure states: First, though the results of measurements on quantum systems generally give random answers, a pure quantum system always provides a definite answer to some question. As a result, a quantum system is "unreadable" (in the sense of providing a definite result of measurement on it) to someone unaware of this question. Second, because a single quantum cannot be cloned (the no-cloning theorem), the system cannot be copied without having been read. Weisner's idea was to create serial numbers for paper currency by embedding in each bill a series of single-photon traps and filling them with a series of linearly polarized photons, each polarization standing for a particular number. If the series were composed of "nonorthogonal" (that is, prepared to answer different questions precisely) polarized photons, say, linearly polarized in both the horizontal/vertical directions and in the diagonal directions, then only the banks, which knew the precise directions to check, would be able to verify the number on the currency. Not having that specialized knowledge, counterfeiters would be unable to read or duplicate it without error. In fact, because measurement collapses the state to the observed result, any counterfeiter's attempt at reading the numbers could be detected by the bank with some probability.

Weisner's idea was ingenious though completely impractical. Yet, in the hands of Weisner's old college friend Bennett and Bennett's collaborator Gilles Brassard, it was transformed into a method for two parties to establish a secret encryption key while not allowing an eavesdropper to go undetected. One party creates a sequence of nonorthogonal photons, each polarized randomly either along the horizontal/vertical direction or

along the diagonals, and sends them, one at a time, to the other party. The receiver can then measure each photon, randomly choosing one of the two bases. Because the sender can predict the measurement result only if the receiver and sender use the same basis, after the measurement the two need to communicate which basis each had used and discard the cases with different bases. Even if eavesdroppers listen to the conversation on a public channel and have access to the photon as it is being transferred, they can neither copy (clone) the photon (so as to store and measure it when its basis is finally announced) nor measure it in a random basis during transmission without affecting its polarization if they choose the wrong basis. The original parties always check the statistics of a small sample of the shared key to see if some process, or an eavesdropper, has affected the photons in flight and then use methods to insure, with high probability, the privacy of the shared key.

The central fact that single quantum systems in an unknown state cannot be cloned, or copied exactly, was proven by Bill Wootters and Wojciech Zurek, in 1982. Their elegant proof uses only the fact that quantum mechanics is a linear theory, in particular, that the principle of superposition always holds (see the box "The No-Cloning Theorem" on page 79). (Dennis Dieks proved the theorem independently that same year.)

Between 1985 and 1994, many people contributed to defining the specific elements of a universal quantum computer, to exploring categories of algorithms that might work more efficiently on a quantum computer, to developing applications of quantum information to communication, and in general, to developing the theory of quantum information in a way that paralleled the theory of classical information. But the interest was mainly confined to a relatively small

group within the research community.

Then, without warning, the field broke wide open. Peter Shor demonstrated that finding the prime factors of an integer, a problem with great practical import, could be solved efficiently on a quantum computer. His solution took advantage of the mathematical fact that the remainders obtained when successive integral powers of any number x were divided by a fixed number N followed a cyclic pattern, and the corresponding period r was directly related to a factor of N . Shor's algorithm arranges an interference between the evaluations of a large sequence of these remainders so as to determine the period of the cycle with small error probability.

It is hard to overestimate how important Shor's work was for converting quantum computing and quantum information from an esoteric field involving only a few specialists to a field of general interest and real funding. One of the central problems in cryptography involves sending an encryption key when no private channel is available. Apart from the quantum key-distribution techniques described earlier, the best available methods in use today rely on the difficulty of factoring products of very large primes. To decrypt information, one has to find a solution to the so-called "discrete logarithm problem," whose practical solution calls for knowing the prime factors of an enormous number (see the box "Public-Key Cryptography: RSA" on page 72). Shor's proof that quantum computers could factor large numbers efficiently means that, if a quantum computer of sufficient power could be built, it would put at risk all such cryptographic methods. And these methods have been widely used to secure banking transactions, exchanges between intelligence agencies, and transactions over the Internet. Given the importance of his work, Shor was awarded the

Nevalinna Prize for mathematical aspects of information science.

Both building a quantum computer and developing new cryptographic protocols such as quantum key distribution took on the aura of urgency. It seemed that these projects were not only interesting but also necessary from the point of view of security. Funding became available for mathematicians to find algorithms other than Shor's that could take advantage of quantum information. The most important one found to date is Grover's algorithm for unstructured searches. Many experimentalists were supported to try implementing what the mathematicians and theoretical physicists said could in principle be done. Ideas for constructing new qubits were cropping up everywhere. And excitement was generated in the popular press. But looming in the background was the certain knowledge that quantum states are fragile. Errors would inevitably occur, for example, through coupling to the environment. One had to find a way of preventing these without destroying the quantum states, which carry the information. That problem was solved in principle by Shor and Andrew Steane. They invented a scheme for error correction analogous to the strategies used for classical information. In 1998, Manny Knill, Raymond Laflamme, and Zurek proved the existence of an error bound, below which a quantum computation of arbitrary size could be implemented to arbitrary accuracy. Independent proofs of related results were done by Dorit Aharonov and Michael Ben-Or, Alexei Kitaev, and John Preskill. Implementing quantum computation in the laboratory became a realistic and compelling goal. Thus began a worldwide effort to build a quantum computer and to explore all the ways in which quantum information could impact science and technology. ■

Acknowledgments

The authors would like to thank Paul Benioff, Manny Knill, Salman Habib, and Wojciech Zurek for helpful comments and suggestions. One of the authors (Necia) would like to thank Tanmoy for sharing his deep understanding of quantum mechanics in this forum.

Further Reading

- Bell, J. 1964. On the Einstein Podolsky Rosen Paradox. *Physics* **1** (3): 195.
- Benioff, P. A. 1982. Quantum Mechanical Models of Turing Machines that Dissipate no Energy. *Phys. Rev. Lett.* **48** (23): 1581.
- Bennett, C. H. 1973. Logical Reversibility of Computation. *IBM J. Res. Dev.* **6**: 525.
- Deutsch, D. 1985. Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer. *Proc. R. Soc. London, Ser. A* **400**: 97.
- Feynman, R. P. 1986. Quantum Mechanical Computers. *Found. Phys.* **16** (6): 507.
- Landauer, R. 1994. Zig-Zag Path to Understanding. In *Proceedings of the Workshop on Physics and Computation PhysComp '94*. Los Alamitos, CA: IEEE Computer Society Press.
- . 1999. Information Is Inevitably Physical. In *Feynman and Computation: Exploring the Limits of Computers*. Edited by A. J. G. Hey. Cambridge: Perseus Books.
- Lloyd, S. 1999. News and Views. *Nature* **400**: 720.
- Wheeler, J. A. 1984. "Bits, Quanta, and Meaning." In *Theoretical Physics Meeting: Commemorative Volume on the Occasion of Eduardo Caianiello's Sixtieth Birthday*. Edited by A. Giovanni, M. Marinaro, F. Mancini, and A. Rimini. Naples, Italy: Edizioni Scientifici Italiani.

Additional references for John Wheeler and Richard Feynman are on pages vi–ix. For papers delivered at the 1981 MIT workshop Physics of Computation, see the *International Journal of Theoretical Physics* **21** (1), 1982.

Tanmoy Bhattacharya
Necia Grant Cooper
September 2002

About This Volume and Quantum Research at Los Alamos

In line with six decades of Laboratory tradition, the breadth of Los Alamos research in quantum science spans the gamut from fundamental questions in quantum theory and measurement to practical applications of quantum information science. The Los Alamos program started in the early 1980s with Wojciech Zurek and his postdoctoral fellows conducting a lively investigation into the emergence of classical reality from the quantum world. Zurek developed the theory of decoherence, which recognizes the role of the coupling between real quantum systems and the environment in the rapid loss of the coherence that endows quantum states with their special properties. In this volume, he surveys the progress in understanding decoherence since his now classic article published in *Physics Today*. In related pieces, Salman Habib and Tanmoy Bhattacharya apply the model of continuous measurement to describe the quantum-to-classical transition and to explore the possibility of controlling quantum systems through continuous quantum feedback. In 1994, as part of the general expansion of interest in quantum computing, Raymond Laflamme and Manny Knill joined Zurek in ground-breaking studies of quantum error correction, which can prevent quantum computers from falling prey to decoherence, and later adapted nuclear magnetic resonance (NMR) technology with molecules to test theoretical ideas in quantum computing.

In the early 1990s, in a parallel development at Los Alamos, Richard Hughes started to implement the quantum cryptographic protocols of Charles Bennett and Gilles Brassard. Hughes, Beth Nordholt, Paul Kwiat, Daniel James, and other colleagues

and postdoctoral fellows gradually expanded their studies of quantum cryptography to include quantum state entanglement of photon pairs and ion-trap quantum computing, in which the qubits are single ions trapped in a linear array inside an electromagnetic trap. In the late 1990s, Chris Hammel started a collaboration with Bruce Kane, Bob Clark, and the quantum technology center in Sydney, Australia, to develop a solid-state quantum computer.

This volume is dedicated to conveying the intellectual excitement of this new field. It opens with an elegant hands-on primer in which Knill and his colleagues define the basic unit of quantum information and introduce all the elements needed to process quantum information. The presentation culminates with a description of a simple quantum network for solving a real problem and a step-by-step solution that shows how the quantum operations produce the answer. The primer ends with a brief but realistic assessment of the advantages of quantum information, particularly for computation. It is a good place to gain a perspective on the future.

Communication, the other major task of information processing, has been profoundly altered by the ideas of quantum information science. Quantum teleportation, quantum cryptography, and other efficient communication schemes exploit the simplest qubit, a linearly polarized photon, to achieve their goals. Often, the use of maximally entangled pairs, or Bell states, has a definite advantage in these contexts. In their article on entanglement, Kwiat and James succeed in explaining and demystifying those schemes. Hughes and Nordholt have developed a working quantum crypto-

graphic system in fiber optics and free space. In their article, they explain both the protocols developed by Bennett and Brassard and their experimental systems in very simple language, accessible to a wide audience.

Most efforts to build a scalable quantum computer struggle with how to construct single qubits and examine their properties. Only ion traps, cavity quantum electrodynamics, and liquid NMR have been used successfully for manipulation of more than one qubit. Laflamme, Knill, and colleagues explain their methods for adapting liquid NMR to a quantum information-processing system. Although the quantum states describing this form of information processing are provably not entangled at any time and the system cannot be scaled up much beyond ten qubits, research at Los Alamos has demonstrated the establishment of well-defined initial states, the system's controlled evolution in the presence of real-world noisy environments, and the ability to read out significant results of a computation from a single qubit.

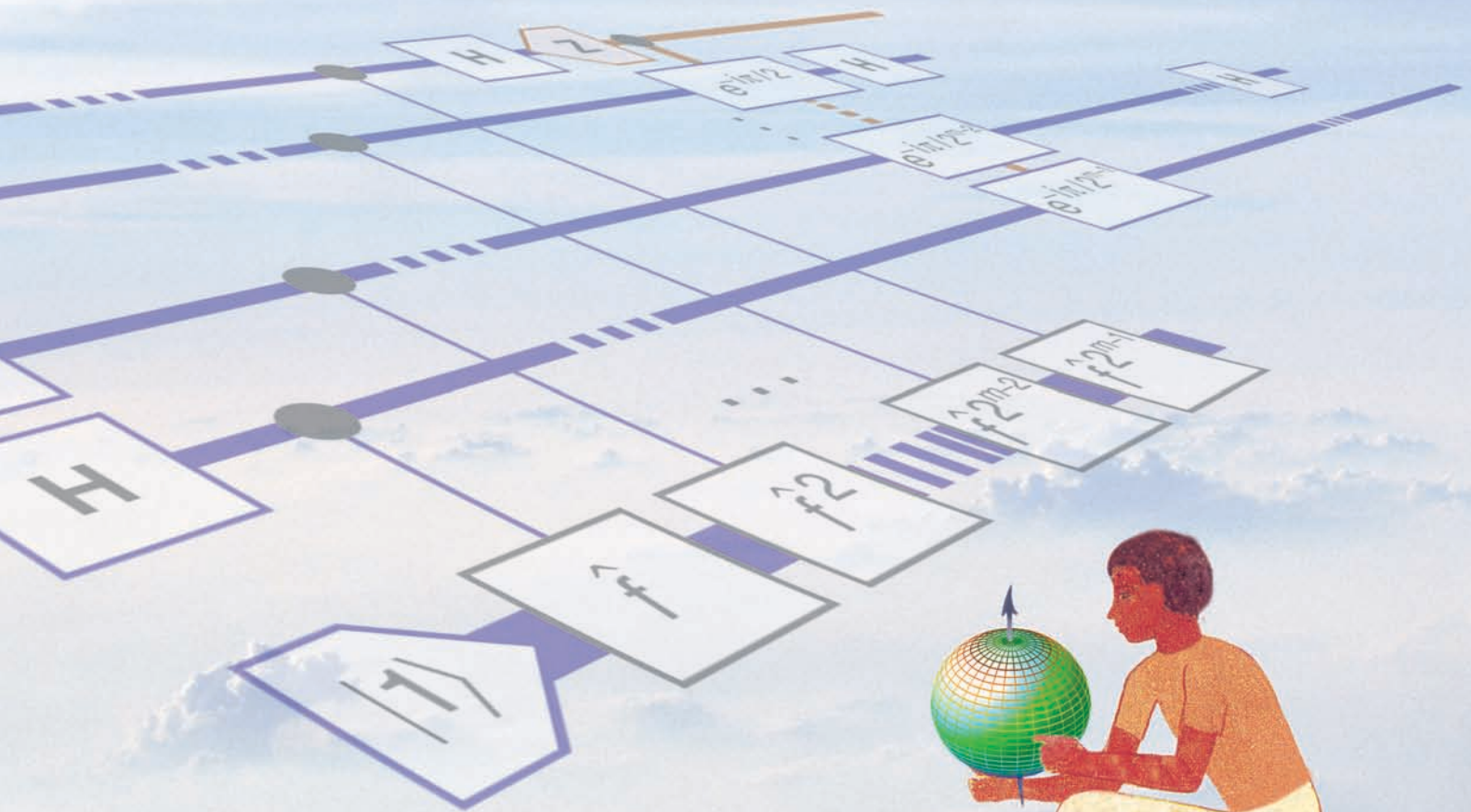
To date, however, quantum information science has far more results from theory than experiment. In his article, Eddy Timmermans explains how dilute Bose-Einstein condensates, many-body quantum states created in atom traps, have become new "laboratories" for studying fundamental quantum phenomena. Dave Vieira and colleagues at Los Alamos are developing an experimental capability in this area.

The diverse quantum efforts at Los Alamos are now supported and fostered by the Quantum Information Research Institute. Contact the steering committee at qsc@lanl.gov for further information. ■

Quantum Information Processing

A hands-on primer

Emanuel Knill, Raymond Laflamme, Howard N. Barnum, Diego A. Dalvit, Jacek J. Dziarmaga, James E. Gubernatis, Leonid Gurvits, Gerardo Ortiz, Lorenza Viola, and Wojciech H. Zurek



$$\alpha|0\rangle + \beta|1\rangle$$

Quantum information processing, Science of—The theoretical, experimental and technological areas covering the use of quantum mechanics for communication and computation
—Kluwer Encyclopedia of Mathematics, Supplement II

Research conducted in the last few decades has established that quantum information, or information based on quantum mechanics, has capabilities that exceed those of traditional “classical” information. For example, in communication, quantum information enables quantum cryptography, which is a method for communicating in secret. Secrecy is guaranteed because eavesdropping attempts necessarily disturb the exchanged quantum information without revealing the content of the communication. In computation, quantum information enables efficient simulation of quantum physics, a task for which general-purpose, efficient, classical algorithms are not known to exist. Quantum information also leads to efficient algorithms for factoring large numbers, which is believed to be difficult for classical computers. An efficient factoring algorithm would break the security of commonly used public-key cryptographic codes used for authenticating and securing Internet communications. Yet another application of quantum information improves the efficiency with which unstructured search problems can be solved. Quantum unstructured search may make it possible to solve significantly larger instances of optimization problems, such as the scheduling and traveling salesman problems.

Because of the capabilities of quantum information, the science of quantum information processing is now a prospering, interdisciplinary field focused on better understanding the possibilities and limitations of the underlying theory, on developing new applications of quantum information, and on physically realizing controllable quantum devices. The purpose of this primer is to provide an elementary introduction to quantum information processing (see Part II), and then to briefly explain how we hope to exploit the advantages of quantum information (see Part III). These two sections can be read independently. For reference, we have included a glossary of the main terms of quantum information (see page 33).

When we use the word “information,” we generally think of the things we can talk about, broadcast, write down, or otherwise record. Such records can exist in many forms, such as sound waves, electrical signals in a telephone wire, characters on paper, pit patterns on an optical disk, or magnetization on a computer hard disk. A crucial property of information is that it is fungible: It can be represented in many different physical forms and easily converted from one form to another without changing its meaning. In this sense, information is independent of the devices used to represent it but requires at least one physical representation in order to be useful.

We call the familiar information stored in today’s computers classical or deterministic to distinguish it from quantum information. It is no accident that classical information is the basis of all human knowledge. Any information passing through our senses is best modeled by classical discrete or continuous information. Therefore, when considering any other kind of information, we need to provide a method for extracting classically meaningful information. We begin by recalling the basic ideas of classical information in a way that illustrates the general procedure for building an information-processing theory.

Part I: Classical Information

The basic unit of classical deterministic information is the bit, an abstract entity or system that can be in one of the two states symbolized by 0 and 1. At this point, the symbols for the two states have no numeric meaning. That is why we have used a font different from that for the numbers 0 and 1. By making a clear distinction between the bit and its states, we emphasize that a bit should be physically realized as a system or device whose states correspond to the ideal bit's states. For example, if you are reading this primer on paper, the system used to realize a bit is a reserved location on the surface, and the state depends on the pattern of ink (0 or 1) in that location. In a computer, the device realizing a bit can be a combination of transistors and other integrated-circuit elements with the state of the bit determined by the distribution of charge.

In order to make use of information, it must be possible to manipulate (or process) the states of information units. The elementary operations that can be used for this purpose are called gates. Two one-bit gates are the **not** and **reset** gates. Applying the **not** gate to a bit has the effect of flipping the state of the bit. For example, if the initial state of the bit is 0, then the state after applying **not** is **not** (0) = 1. We can present the effect of the gate in the following form:

Initial State		Final State
0	→	not (0) = 1 , and
1	→	not (1) = 0 .

(1)

The **reset** gate sets the state to 0 regardless of the input:

Initial State		Final State
0	→	reset (0) = 0 , and
1	→	reset (1) = 0 .

(2)

By applying a combination of **not** and **reset** gates, one can transform the state of a bit in every possible way.

Information units can be combined to represent more information. Bits are typically combined into sequences. The states of such a sequence are symbolized by strings of state symbols for the constituent bits. For example, a two-bit sequence can be in one of the following four states: 00, 01, 10, and 11. The different bits are distinguished by their position in the sequence.

The one-bit gates can be applied to any bit in a sequence. For example, the **not** gate applied to the second bit of a three-bit sequence in the state 011 changes the state to 001.

One-bit gates act independently on each bit. To compute with multiple bits, we need gates whose action can correlate the states of two or more bits. One such gate is the **nand** (“not and”) gate, which acts on two bits in a bit sequence. Its effect is to set the state of the first bit to 0 if both the first and the second bit are 1; otherwise, it sets it to 1. Here is what happens when **nand** is applied to two consecutive bits:

Initial State		Final State
00	→	nand (00) = 10 ,
01	→	nand (01) = 11 ,
10	→	nand (10) = 10 , and
11	→	nand (11) = 01 .

(3)

The **nand** gate can be applied to any two bits in a sequence. For example, it can be applied to the fourth and second bits (in this order) of four bits, in which case the initial state 1101 is transformed to 1100, setting the fourth bit to 0.

Other operations on bit sequences include adding a new bit to the beginning (prepend) or end (append) of a sequence. The new bit is always initialized to 0. It is also possible to discard the first or last bit regardless of its state. Versions of these operations that are conditional on the state of another bit may also be used. An example is the conditional append operation: “If the k^{th} bit is in the state 0, then append a bit.”

The gates just introduced suffice for implementing arbitrary state transformations of a given bit sequence. Instructions for applying gates in a particular order are called a circuit. An important part of investigations in information processing is to determine the minimum resources required to perform information-processing tasks. For a given circuit, the two primary resources are the number of gates and the total number of bits used. The circuit complexity of a desired transformation is the minimum number of gates needed to implement it.

The model of computation defined by the ability to apply gates in a fixed sequence is called the circuit model. Classical computation extends the circuit model by providing a means for repeating blocks of instructions indefinitely or until a desired condition is achieved. In principle, it is possible to conceive of a general-purpose computer as a device that repeatedly applies the same circuit to the beginnings of several bit sequences. In this article, we take for granted a traditional programmable computer based on classical information. Thus, a quantum algorithm is a program written for such a computer with additional instructions for applying gates to quantum information. The computational power of this model is equivalent to that of other general-purpose models of quantum computation, such as quantum Turing machines (Yao 1993).

For an introduction to algorithms and their analysis, refer to Thomas Cormen et al. (1990). Christos Papadimitriou wrote (1994) a useful textbook on computational complexity with an introduction to classical computation and computational machine models.

Part II: Quantum Information

The foundations of an information-processing theory can be constructed by the procedure we followed in the previous section:

1. Define the basic unit of information.
2. Give the means for processing one unit.
3. Describe how multiple units can be combined.
4. Give the means for processing multiple units.
5. Show how to convert the content of any of the extant units to classical information.

Note that the last step was not required for classical information processing.

In this section, we follow the general procedure for defining an information-processing theory to introduce quantum information processing. A simple example that exhibits the advantages of quantum information is given in the section “The Parity Problem” on page 21. A version of the quantum factoring algorithm is described immediately following this article in “From Factoring to Phase Estimation” on page 38.

The Quantum Bit

The fundamental resource and basic unit of quantum information is the quantum bit (qubit), which behaves like a classical bit enhanced by the superposition principle (see discussion in this section). From a physical point of view, a qubit is represented by an ideal two-state quantum system. Examples of such systems include photons (vertical and horizontal polarization), electrons and other spin-1/2 systems (spin-up and -down), and systems defined by two energy levels of atoms or ions. From the beginning, the two-state system played a central role in studies of quantum mechanics. It is the simplest quantum system, and in principle, all other quantum systems can be modeled in the state space of collections of qubits.

From the information-processing point of view, a qubit’s state space contains the two “logical,” or computational, states $|0\rangle$ and $|1\rangle$. The so-called “ket” notation for these states was introduced by Paul Dirac, and its variations are widely used in quantum physics. One can think of the pair of symbols $|$ and \rangle as representing the qubit system. Their content specifies a state for the system. In this context, 0 and 1 are system-independent state labels. When, say, 0 is placed within the ket, the resulting expression $|0\rangle$ represents the corresponding state of a specific qubit.

The initial state of a qubit is always one of the logical states. Using operations to be introduced later, we can obtain states that are superpositions of the logical states. Superpositions can be expressed as sums $\alpha|0\rangle + \beta|1\rangle$ over the logical states with complex coefficients. The complex numbers α and β are the amplitudes of the superposition. The existence of such superpositions of distinguishable states of quantum systems is one of the basic tenets of quantum theory and is called the superposition principle. Another way of writing a general superposition is as a vector:

$$\alpha|0\rangle + \beta|1\rangle \leftrightarrow \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad (4)$$

where the two-sided arrow is used to denote the correspondence between expressions that mean the same thing.

The qubit states that are superpositions of the logical states are called pure states: A superposition $\alpha|0\rangle + \beta|1\rangle$ is a pure state if the corresponding vector has length 1, that

is, $|\alpha|^2 + |\beta|^2 = 1$. Such a superposition or vector is said to be normalized. (For a complex number given by $\gamma = x + iy$, one can evaluate $|\gamma|^2 = x^2 + y^2$. Here, x and y are the real and imaginary part of γ , and the symbol i is a square root of -1 , that is, $i^2 = -1$. The conjugate of γ is $\bar{\gamma} = x - iy$. Thus, $|\bar{\gamma}|^2 = \gamma\bar{\gamma}$). Here are a few examples of states given in both the ket and vector notation:

$$|\psi_1\rangle = \frac{(|0\rangle + |1\rangle)}{\sqrt{2}} \leftrightarrow \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}, \quad (5)$$

$$|\psi_2\rangle = \frac{3}{5}|0\rangle - \frac{4}{5}|1\rangle \leftrightarrow \begin{pmatrix} 3/5 \\ -4/5 \end{pmatrix}, \quad \text{and} \quad (6)$$

$$|\psi_3\rangle = \frac{i3}{5}|0\rangle - \frac{i4}{5}|1\rangle \leftrightarrow \begin{pmatrix} i3/5 \\ -i4/5 \end{pmatrix}. \quad (7)$$

The state $|\psi_3\rangle$ is obtained from $|\psi_2\rangle$ by multiplication with i . It turns out that two states cannot be distinguished if one of them is obtained by multiplying the other by a phase $e^{i\theta}$. Note how we have generalized the ket notation by introducing expressions such as $|\psi\rangle$ for arbitrary states.

The superposition principle for quantum information means that we can have states that are sums of logical states with complex coefficients. There is another, more familiar type of information, whose states are combinations of logical states. The basic unit of this type of information is the probabilistic bit (pbit). Intuitively, a pbit can be thought of as representing the as-yet-undetermined outcome of a coin flip. Since we need the idea of probability to understand how quantum information converts to classical information, we briefly introduce pbits.

A pbit's state space is a probability distribution over the states of a bit. One very explicit way to symbolize such a state is by using the expression $\{p:0, (1-p):1\}$, which means that the pbit has probability p of being 0 and $1-p$ of being 1. Thus, a state of a pbit is a probabilistic combination of the two logical states, where the coefficients are nonnegative real numbers summing to 1. A typical example is the unbiased coin in the process of being flipped. If tail and head represent 0 and 1, respectively, the coin's state is $\{1/2:0, 1/2:1\}$. After the outcome of the flip is known, the state collapses to one of the logical states 0 and 1. In this way, a pbit is converted to a classical bit. If the pbit is probabilistically correlated with other pbits, the collapse associated with learning the pbit's logical state changes the overall probability distribution by a process called conditioning on the outcome.

A consequence of the conditioning process is that we never actually "see" a probability distribution. We only see classical deterministic bit states. According to the frequency interpretation of probabilities, the original probability distribution can only be inferred after one looks at many independent pbits in the same state $\{p:0, (1-p):1\}$: In the limit of infinitely many pbits, p is given by the fraction of pbits seen to be in the state 0. As we will explain, we can never see a general qubit state either. For qubits, there is a process analogous to conditioning. It is called measurement and converts qubit states to classical information.

Information processing with pbits has many advantages over deterministic information processing with bits. One advantage is that algorithms are often much easier to design and

analyze if they are probabilistic. Examples include many optimization and physics simulation algorithms. In some cases, the best available probabilistic algorithm is more efficient than any known deterministic algorithm. An example is an algorithm for determining whether a number is prime or not. It is not known whether every probabilistic algorithm can be derandomized efficiently. There are important communication problems that can be solved probabilistically but not deterministically. For a survey of these algorithms, see Rajiv Gupta (1994a).

What is the difference between bits, pbits, and qubits? One way to visualize the difference and see the enrichment provided by pbits and qubits is shown in Figure 1.

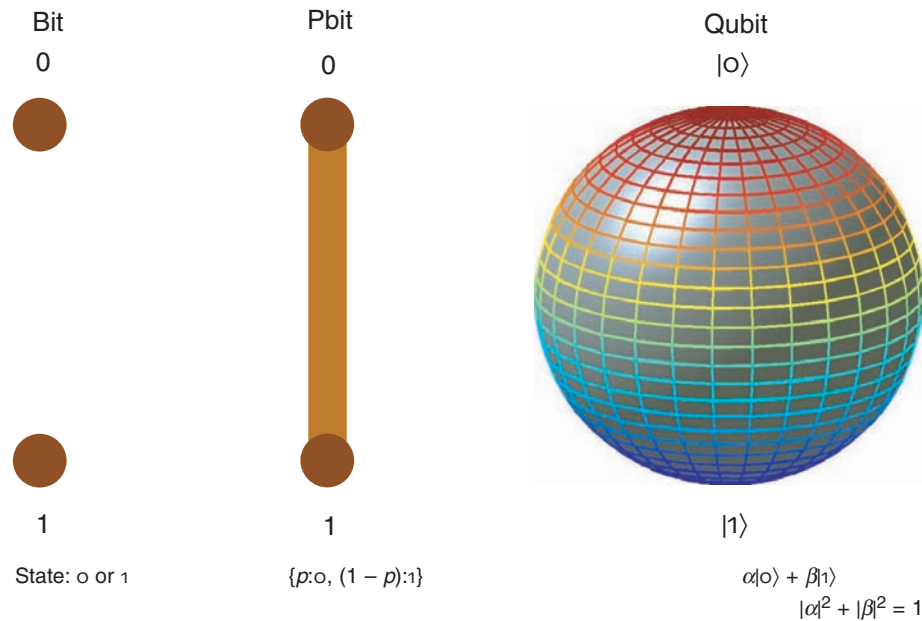


Figure 1. Comparing State Spaces of Different Information Units

The states of a bit correspond to two points. The states of a pbit can be thought of as convex combinations of a bit's states and therefore can be visualized as lying on the line connecting the two bit states. A qubit's pure states correspond to the surface of the unit sphere in three dimensions, where the logical states correspond to the poles. This representation of qubit states is called the Bloch sphere. The explicit correspondence is discussed at the end of the section "Mixtures and Density Operators." Also refer to the definition and use of the Bloch sphere in the article "NMR and Quantum Information Processing" on page 226. There, the correspondence between the pure states and the sphere is physically motivated and comes from a way of viewing a spin-1/2 system as a small quantum magnet. Intuitively, a state is determined by the direction of the north pole of the magnet.

Processing One Qubit

The quantum version of the **not** gate for bits exchanges the two logical states; that is, using ket notation,

$$\text{not}(\alpha|0\rangle + \beta|1\rangle) = \alpha|1\rangle + \beta|0\rangle = \beta|0\rangle + \alpha|1\rangle. \tag{8}$$

In vector notation, this equation becomes

$$\mathbf{not} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \beta \\ \alpha \end{pmatrix}. \quad (9)$$

Another way of expressing the effect of **not** is by multiplying the vector by a matrix representing **not**,

$$\mathbf{not} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \beta \\ \alpha \end{pmatrix}, \quad (10)$$

so that we can identify the action of **not** with the matrix

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

An even simpler gate is the one that does nothing. We call it the **noop** gate, and its matrix form is the identity matrix, as shown in the following equation:

$$\mathbf{noop} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (11)$$

The **noop** and **not** gates are reversible. In other words, we can undo their actions by applying other gates. For example, the action of the **not** gate can be undone by another **not** gate. The action of every reversible quantum gate can be represented by matrix multiplication, where the matrix has the additional property of preserving vector lengths. Such matrices are called unitary and are characterized by the equation $A^\dagger A = \mathbb{1}$, where A^\dagger is the conjugate transpose of A and $\mathbb{1}$ is the identity matrix. (The conjugate transpose of a matrix is computed by flipping that matrix across the main diagonal and conjugating the complex numbers). For gates represented by a matrices, the unitarity condition is necessary and sufficient for ensuring that pure states get mapped to pure states.

Because qubit states can be represented as points on a sphere, reversible one-qubit gates can be thought of as rotations of the Bloch sphere. This is why such quantum gates are often called rotations. As explained in detail on page 232 in the article “NMR and Quantum Information Processing”, rotations around the x -, y -, and z -axis are in a sense generated by the three Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (12)$$

each of which represents a one-qubit gate. For example, a rotation around the x -axis by an angle ϕ is given by $e^{-i\sigma_x\phi/2} = \cos(\phi/2)\mathbb{1} - i\sin(\phi/2)\sigma_x$. To obtain this identity, one can use the power series for e^A , $e^A = \sum_{k=0}^{\infty} (1/k!)A^k$, and exploit the fact that $\sigma_x^2 = \mathbb{1}$ to simplify the expression. Here are some gates that can be defined with the help of rotations:

$$90^\circ \text{ } x\text{-rotation: } \mathbf{rotx}_{90^\circ} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix},$$

$$90^\circ \text{ } y\text{-rotation: } \mathbf{roty}_{90^\circ} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix},$$

$$\phi \text{ } z\text{-rotation: } \mathbf{rotz}_\phi = \begin{pmatrix} e^{-i\phi/2} & 0 \\ 0 & e^{i\phi/2} \end{pmatrix}, \text{ and}$$

$$\text{Hadamard gate: } \mathbf{H} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{13}$$

The rotation gates often show up in controlling spins or ions with radio-frequency pulses or lasers. The Hadamard gate is used primarily by quantum programmers. It can be expressed as a product of a 90° y -rotation and σ_z .

To check directly that the rotation gates are reversible, one can determine their inverses. In this case and as expected, the inverse of a rotation is the rotation around the same axis in the opposite direction. For example, the inverses of the \mathbf{roty}_{90° and \mathbf{rotz}_ϕ gates are given by

$$\mathbf{roty}_{-90^\circ} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \text{ and } \mathbf{rotz}_{-\phi} = \begin{pmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{pmatrix}. \tag{14}$$

Another useful property of the rotation gates is that the angles add when rotations are applied around the same axis. For example, $\mathbf{rotz}_\phi \mathbf{rotz}_\theta = \mathbf{rotz}_{\phi+\theta}$.

The Bra-Ket Notation for Logic Gates. The ket notation can be extended so that we can write gates in a compact form that readily generalizes to multiple qubits. To do so, we have to introduce expressions such as $\langle \psi | = \alpha \langle 0 | + \beta \langle 1 |$. This is called the “bra” notation. The terminology comes from the term “bracket:” The bra is the left, and the ket is the right part of a matched pair of brackets. From the vector point of view, $\langle \psi |$ corresponds to the row vector (α, β) . Note that a column vector multiplied by a row vector yields a matrix. In the bra-ket notation, this corresponds to multiplying a ket $|\psi\rangle$ by a bra $\langle \phi |$, written as $|\psi\rangle\langle \phi |$. Since this represents an operator on states, we expect to be able to compute the effect of $|\psi\rangle\langle \phi |$ on a state $|\phi\rangle$ by forming the product. To be able to evaluate such products with one-qubit bras and kets, we need the following two rules: distributivity and inner-product evaluation.

Distributivity

You can rewrite sums and products using distributivity. For example,

$$\left(\frac{3}{5} \langle 0 | + \frac{4}{5} \langle 1 | \right) i | 1 \rangle = \frac{i3}{5} \langle 0 | 1 \rangle + \frac{i4}{5} \langle 1 | 1 \rangle. \tag{15}$$

Observe that we can combine the amplitudes of terms, but we cannot rearrange the order of the bras and kets in a product.

Inner-Product Evaluation

The product of a logical bra and a logical ket is evaluated according to the identities

$$\begin{aligned}\langle 0|0\rangle &= 1 \\ \langle 0|1\rangle &= 0, \\ \langle 1|0\rangle &= 0, \text{ and} \\ \langle 1|1\rangle &= 1.\end{aligned}\tag{16}$$

It follows that for logical states, if a bra multiplies a ket, the result cancels unless the states match, in which case the answer is 1. Applying inner-product evaluation to Equation (15) results in

$$\frac{i3}{5}\langle 0|1\rangle + \frac{i4}{5}\langle 1|1\rangle = \frac{i3}{5}0 + \frac{i4}{5}1 = \frac{i4}{5}.\tag{17}$$

To simplify the notation, we can omit one of the two vertical bars in products such as $\langle a|b\rangle$ and write $\langle a|b\rangle$.

To understand inner-product evaluation, think of the expressions as products of row and column vectors. For example,

$$\langle 0|1\rangle \leftrightarrow \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0.\tag{18}$$

That is, as vectors, the two states $|0\rangle$ and $|1\rangle$ are orthogonal. In general, if $|\phi\rangle$ and $|\psi\rangle$ are states, then $\langle\phi|\psi\rangle$ is the inner product, or “overlap,” of the two states. In the expression for the overlap, we compute $\langle\phi|$ from $|\phi\rangle = \bar{\alpha}|0\rangle + \bar{\beta}|1\rangle$ by conjugating the coefficients and converting the logical kets to bras: $\langle\phi| = \alpha\langle 0| + \beta\langle 1|$. In the vector representation, this is the conjugate transpose of the column vector for $|\phi\rangle$, so the inner product agrees with the usual one. Two states are orthogonal if their overlap is zero. We write $|\phi\rangle^\dagger = \langle\phi|$ and $\langle\phi|^\dagger = |\phi\rangle$.

Every linear operator on states can be expressed with the bra-ket notation. For example, the bra-ket expression for the **noop** gate is **noop** = $|0\rangle\langle 0| + |1\rangle\langle 1|$. To apply **noop** to a qubit, you multiply its state on the left by the bra-ket expression

$$\begin{aligned}\text{noop}(\alpha|0\rangle + \beta|1\rangle) &= (|0\rangle\langle 0| + |1\rangle\langle 1|)(\alpha|0\rangle + \beta|1\rangle) \\ &= |0\rangle\langle 0|(\alpha|0\rangle + \beta|1\rangle) + |1\rangle\langle 1|(\alpha|0\rangle + \beta|1\rangle) \\ &= \alpha|0\rangle\langle 0|0\rangle + \beta|0\rangle\langle 0|1\rangle + \alpha|1\rangle\langle 1|0\rangle + \beta|1\rangle\langle 1|1\rangle \\ &= \alpha|0\rangle 1 + \beta|0\rangle 0 + \alpha|1\rangle 0 + \beta|1\rangle 1 \\ &= \alpha|0\rangle + \beta|1\rangle.\end{aligned}\tag{19}$$

One way to think about an operator such as $|a\rangle\langle b|$ is to notice that, when it is used to operate on a ket expression, the $\langle b|$ picks out the matching kets in the state, which are

then changed to $|a\rangle$. For example, we can write the **not** operation as $\mathbf{not} = |0\rangle\langle 1| + |1\rangle\langle 0|$.

The coefficients of the $|a\rangle\langle b|$ in a bra-ket representation of a gate correspond to matrix entries in the matrix representation. The relationship is defined by

$$\alpha_{00}|0\rangle\langle 0| + \alpha_{01}|0\rangle\langle 1| + \alpha_{10}|1\rangle\langle 0| + \alpha_{11}|1\rangle\langle 1| \leftrightarrow \begin{pmatrix} \alpha_{00} & \alpha_{01} \\ \alpha_{10} & \alpha_{11} \end{pmatrix}. \tag{20}$$

Two Quantum Bits

Some states of two quantum bits can be symbolized by the juxtaposition (or multiplication) of the states of each quantum bit. In particular, the four logical states $|0\rangle|0\rangle$, $|0\rangle|1\rangle$, $|1\rangle|0\rangle$, and $|1\rangle|1\rangle$ are acceptable pure states for two quantum bits. In these expressions, we have distinguished the qubits by position (first or second). It is easier to manipulate state expressions if we explicitly name the qubits, say, A and B. We can then distinguish the kets by writing, for example, $|\psi\rangle_A$ for a state of qubit A. Now, the state $|0\rangle|1\rangle$ can be written with explicit qubit names (or labels) as

$$|0\rangle_A |1\rangle_B = |1\rangle_B |0\rangle_A = |01\rangle_{AB} = |10\rangle_{BA}. \tag{21}$$

Having explicit labels allows us to unambiguously reorder the states in a product of states belonging to different qubits. We say that kets for different qubits “commute.”

So far, we have seen four states of two qubits, which are the logical states that correspond to the states of two bits. As in the case of one qubit, we can use the superposition principle to get all the other pure states. Each state of two qubits is therefore of the form

$$\alpha|00\rangle_{AB} + \beta|01\rangle_{AB} + \gamma|10\rangle_{AB} + \delta|11\rangle_{AB}, \tag{22}$$

where α , β , γ , and δ are complex numbers. Again, there is a column vector form for the state,

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix}, \tag{23}$$

and this vector has to be of unit length, that is, $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$. When using the vector form for qubit states, one has to be careful about the convention used for ordering the coefficients.

Other examples of two-qubit states in ket notation are the following:

$$\begin{aligned}
|\psi_1\rangle_{AB} &= \frac{1}{\sqrt{2}}(|0\rangle_A + |1\rangle_A)|1\rangle_B . \\
|\psi_2\rangle_{AB} &= \frac{1}{\sqrt{2}}(|0\rangle_A - |1\rangle_A) \frac{1}{\sqrt{2}}(|0\rangle_B + i|1\rangle_B) \\
&= \frac{1}{2}(|00\rangle_{AB} + i|01\rangle_{AB} - |10\rangle_{AB} - i|11\rangle_{AB}) . \\
|\psi_3\rangle_{AB} &= \frac{1}{\sqrt{2}}(|00\rangle_{AB} + |11\rangle_{AB}) . \\
|\psi_4\rangle_{AB} &= \frac{1}{\sqrt{2}}(|01\rangle_{AB} - |10\rangle_{AB}) . \tag{24}
\end{aligned}$$

The first two of these states have the special property that they can be written as a product $|\phi_1\rangle_A|\phi_2\rangle_B$ of a state of qubit A and a state of qubit B. The second expression for $|\psi_2\rangle$ shows that the product decomposition is not always easy to see. Such states are called product states. The last two states, $|\psi_3\rangle_{AB}$ and $|\psi_4\rangle_{AB}$, are two of the famous Bell states. They have no such representation as a product of independent states of each qubit. They are said to be entangled because they contain a uniquely quantum correlation between the two qubit systems. Pbits can also have correlations that cannot be decomposed into product states, but the entangled states have additional properties that make them very useful. For example, if Alice and Bob each have one of the qubits that together are in the state $|\psi_3\rangle_{AB}$, they can use them to create a secret bit for encrypting their digital communications (see the article “Quantum State Entanglement” on page 52).

Processing Two Qubits

The simplest way of modifying the state of two qubits is to apply one of the one-qubit gates. If the gates are expressed in the bra-ket notation, all we need to do is add qubit labels so that we know which qubit each bra or ket belongs to. For example, the **not** gate for qubit B is written as

$$\mathbf{not}^{(B)} = |0\rangle_B^B \langle 1| + |1\rangle_B^B \langle 0| . \tag{25}$$

The labels for bra expressions occur as left superscripts. To apply expressions like this to states, we need one more rule, namely, commutation.

Commutation

Kets and bras with different labels can be interchanged in products (they commute). This property is demonstrated by the following example:

$$\begin{aligned}
(|0\rangle_B^B \langle 1|) |01\rangle_{AB} &= |0\rangle_B^B \langle 1| |0\rangle_A |1\rangle_B \\
&= |0\rangle_A |0\rangle_B^B \langle 1| |1\rangle_B \\
&= |0\rangle_A |0\rangle_B^B \langle 1| |1\rangle_B \\
&= |0\rangle_A |0\rangle_B = |00\rangle_{AB} . \tag{26}
\end{aligned}$$

Note that we cannot merge the two vertical bars in expressions such as ${}^B\langle 1|0\rangle_A$ because the two terms belong to different qubits. The bars can only be merged when the expression is an inner product, which requires that the two terms belong to the same qubit.

With the rules for bra-ket expressions in hand, we can apply the **not** gate to one of our Bell states to see how it acts:

$$\begin{aligned}
 \mathbf{not}^{(B)} \frac{1}{\sqrt{2}} (|00\rangle_{AB} + |11\rangle_{AB}) &= (|0\rangle_B^B \langle 1| + |1\rangle_B^B \langle 0|) \frac{1}{\sqrt{2}} (|00\rangle_{AB} + |11\rangle_{AB}) \\
 &= \frac{1}{\sqrt{2}} \left(|0\rangle_B^B \langle 1| (|00\rangle_{AB} + |11\rangle_{AB}) + |1\rangle_B^B \langle 0| (|00\rangle_{AB} + |11\rangle_{AB}) \right) \\
 &= \frac{1}{\sqrt{2}} (|0\rangle_B^B \langle 1| |00\rangle_{AB} + |0\rangle_B^B \langle 1| |11\rangle_{AB} + |1\rangle_B^B \langle 0| |00\rangle_{AB} + |1\rangle_B^B \langle 0| |11\rangle_{AB}) \\
 &= \frac{1}{\sqrt{2}} (|0\rangle_A |0\rangle_B^B \langle 1| |0\rangle_B + |1\rangle_A |0\rangle_B^B \langle 1| |1\rangle_B + |0\rangle_A |1\rangle_B^B \langle 0| |0\rangle_B + |1\rangle_A |1\rangle_B^B \langle 0| |1\rangle_B) \\
 &= \frac{1}{\sqrt{2}} (|0\rangle_A |0\rangle_B 0 + |1\rangle_A |0\rangle_B 1 + |0\rangle_A |1\rangle_B 1 + |1\rangle_A |1\rangle_B 0) \\
 &= \frac{1}{\sqrt{2}} (|1\rangle_A |0\rangle_B + |0\rangle_A |1\rangle_B) = \frac{1}{\sqrt{2}} (|01\rangle_{AB} + |10\rangle_{AB}) . \tag{27}
 \end{aligned}$$

The effect of the gate was to flip the state symbols for qubit B, which results in another Bell state.

The gate $\mathbf{not}^{(B)}$ can also be written as a 4×4 matrix acting on the vector representation of a two-qubit state. However, the relationship between this matrix and the one-qubit matrix is not as obvious as for the bra-ket expression. The matrix is

$$\mathbf{not}^{(B)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \tag{28}$$

which swaps the top two and bottom two entries of a state vector.

One way to see the relationship between the one- and two-qubit representations of the gate $\mathbf{not}^{(B)}$ is to notice that because the **noop** gate acts as the identity and because we can act on different qubits independently, $\mathbf{noop}^{(A)} \mathbf{not}^{(B)} \equiv \mathbf{not}^{(B)}$. The matrix for $\mathbf{not}^{(B)}$ can be expressed as a Kronecker product (\otimes) of the matrices for **noop** and **not**:

$$\begin{aligned}
\text{noop}^{(A)}\text{not}^{(B)} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & 0 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ 0 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & 1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{pmatrix} \\
&= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.
\end{aligned} \tag{29}$$

The Kronecker product of two matrices expands the first matrix by multiplying each entry by the second matrix. A disadvantage of the matrix representation of quantum gates is that it depends on the number and order of the qubits. However, it is often easier to visualize what the operation does by writing down the corresponding matrix.

One cannot do much with one-bit classical gates. Similarly, the utility of one-qubit gates is limited. In particular, it is not possible to obtain a Bell state starting from $|\text{OO}\rangle_{AB}$ or any other product state. We therefore need to introduce at least one two-qubit gate not expressible as the product of two one-qubit gates. The best-known such gate is the controlled-not (**cnot**) gate. Its action can be described by the statement, “if the first bit is 1, flip the second bit; otherwise, do nothing.” The bra-ket and matrix representations for this action are

$$\begin{aligned}
\text{cnot}^{(AB)} &= |\text{O}\rangle_A^A \langle \text{O}| + |1\rangle_A^A \langle 1| \left(|\text{O}\rangle_B^B \langle 1| + |1\rangle_B^B \langle \text{O}| \right) \\
&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.
\end{aligned} \tag{30}$$

The **cnot** gate is reversible because its action is undone if a second **cnot** is applied. This outcome is easy to see by computing the square of the matrix for **cnot**, which yields the identity matrix. As an exercise in manipulating bras and kets, let us calculate the product of two **cnot** gates by using the bra-ket representation:

$$\begin{aligned}
\text{cnot}^{(AB)}\text{cnot}^{(AB)} &= \left(|\text{O}\rangle_A^A \langle \text{O}| + |1\rangle_A^A \langle 1| \left(|\text{O}\rangle_B^B \langle 1| + |1\rangle_B^B \langle \text{O}| \right) \right) \\
&\quad \times \left(|\text{O}\rangle_A^A \langle \text{O}| + |1\rangle_A^A \langle 1| \left(|\text{O}\rangle_B^B \langle 1| + |1\rangle_B^B \langle \text{O}| \right) \right).
\end{aligned} \tag{31}$$

The first step is to expand this expression by multiplying out. Expressions such as $|\mathbf{0}\rangle_A^A \langle \mathbf{0} | | \mathbf{1}\rangle_A^A \langle \mathbf{1} |$ cancel because of the inner-product evaluation rule $\langle \mathbf{0} | \mathbf{1}\rangle_A^A = 0$. One can also reorder bras and kets with different labels and rewrite $|\mathbf{0}\rangle_A^A \langle \mathbf{0} | | \mathbf{0}\rangle_A^A \langle \mathbf{0} | = |\mathbf{0}\rangle_A^A \langle \mathbf{0} |$ to get

$$\begin{aligned}
 \mathbf{cnot}^{(AB)} \mathbf{cnot}^{(AB)} &= |\mathbf{0}\rangle_A^A \langle \mathbf{0} | + |\mathbf{1}\rangle_A^A \langle \mathbf{1} | \left(|\mathbf{0}\rangle_B^B \langle \mathbf{1} | + |\mathbf{1}\rangle_B^B \langle \mathbf{0} | \right) \left(|\mathbf{0}\rangle_B^B \langle \mathbf{1} | + |\mathbf{1}\rangle_B^B \langle \mathbf{0} | \right) \\
 &= |\mathbf{0}\rangle_A^A \langle \mathbf{0} | + |\mathbf{1}\rangle_A^A \langle \mathbf{1} | \left(|\mathbf{0}\rangle_B^B \langle \mathbf{0} | + |\mathbf{1}\rangle_B^B \langle \mathbf{1} | \right) \\
 &= |\mathbf{0}\rangle_A^A \langle \mathbf{0} | + |\mathbf{1}\rangle_A^A \langle \mathbf{1} | \mathbf{noop}^{(B)} \\
 &\cong |\mathbf{0}\rangle_A^A \langle \mathbf{0} | + |\mathbf{1}\rangle_A^A \langle \mathbf{1} | \\
 &= \mathbf{noop}^{(A)} \\
 &\cong 1 .
 \end{aligned} \tag{32}$$

Here we used the fact that, when the bra-ket expression for **noop** is applied to the ket expression for a state, it acts the same as (here denoted by the symbol \cong) multiplication by the number 1.

Using Many Quantum Bits

To use more than two, say, five qubits, we can just start with the state $|\mathbf{0}\rangle_A^A |\mathbf{0}\rangle_B^B |\mathbf{0}\rangle_C^C |\mathbf{0}\rangle_D^D |\mathbf{0}\rangle_E^E$ and apply gates to any one or two of these qubits. For example, $\mathbf{cnot}^{(DB)}$ applies the **cnot** operation from qubit D to qubit B. Note that the order of D and B in the label for the **cnot** operation matters. In the bra-ket notation, we simply multiply the state with the bra-ket form of $\mathbf{cnot}^{(DB)}$ from the left. One can express everything in terms of matrices and vectors, but now the vectors have length $2^5 = 32$, and the Kronecker product expression for $\mathbf{cnot}^{(DB)}$ requires some reordering to enable inserting the operation so as to act on the intended qubits. Nevertheless, to analyze the properties of all reversible (that is, unitary) operations on these qubits, it is helpful to think of the matrices because a lot of useful properties about unitary matrices are known. One important result from this analysis is that every matrix that represents a reversible operation on quantum states can be expressed as a product of the one- and two-qubit gates introduced so far. We say that this set of gates is universal.

For general-purpose computation, it is necessary to have access to arbitrarily many qubits. Instead of assuming that there are infinitely many from the start, it is convenient to have an operation to add a new qubit, namely, **add**. To add a new qubit labeled X in the state $|\mathbf{0}\rangle_X^X$, apply $\mathbf{add}^{(X)}$ to the current state. This operation can only be used if there is not already a qubit labeled X. To implement the $\mathbf{add}^{(X)}$ operation in the bra-ket notation, we multiply the ket expression for the current state by $|\mathbf{0}\rangle_X^X$.

Qubit Measurements

In order to classically access information about the state of qubits, we use the measurement operation **meas**. This is an intrinsically probabilistic process that can be applied to any extant qubit. For information processing, one can think of **meas** as a subroutine or function whose output is either 0 or 1. The output is called the measurement outcome. The probabilities of the measurement outcomes are determined by the current state. The state of the qubit being measured is collapsed to the logical state corresponding to the outcome. Suppose we have just one qubit, currently in the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. Measurement of this qubit has the effect

$$\mathbf{meas}(\alpha|0\rangle + \beta|1\rangle) = \begin{cases} 0: |0\rangle & \text{with probability } |\alpha|^2 \\ 1: |1\rangle & \text{with probability } |\beta|^2 \end{cases} . \quad (33)$$

The classical output is given before the new state for each possible outcome. This measurement behavior explains why the amplitudes have to define unit length vectors: Up to a phase, they are associated with square roots of probabilities.

For two qubits, the process is more involved. Because of possible correlations between the two qubits, the measurement affects the state of the other one too, similar to conditioning for pbits after one “looks” at one of them. As an example, consider the state

$$|\psi\rangle_{AB} = \frac{2}{3}|01\rangle_{AB} + \frac{i2}{3}|10\rangle_{AB} + \frac{1}{3}|00\rangle_{AB} . \quad (34)$$

To figure out what happens when we measure qubit A, we first rewrite the current state in the form $\alpha|0\rangle_A|\phi_0\rangle_B + \beta|1\rangle_A|\phi_1\rangle_B$, where $|\phi_0\rangle_B$ and $|\phi_1\rangle_B$ are pure states for qubit B. It is always possible to do that. For the example given in Equation (34),

$$\begin{aligned} |\psi\rangle_{AB} &= \frac{2}{3}|0\rangle_A|1\rangle_B + \frac{1}{3}|0\rangle_A|0\rangle_B + \frac{i2}{3}|1\rangle_A|0\rangle_B \\ &= |0\rangle_A \left(\frac{2}{3}|1\rangle_B + \frac{1}{3}|0\rangle_B \right) + |1\rangle_A \frac{i2}{3}|0\rangle_B \\ &= \frac{\sqrt{5}}{3}|0\rangle_A \left(\frac{2}{\sqrt{5}}|1\rangle_B + \frac{1}{\sqrt{5}}|0\rangle_B \right) + \frac{i2}{3}|1\rangle_A|0\rangle_B , \end{aligned} \quad (35)$$

$$\text{so } \alpha = \frac{\sqrt{5}}{3} , \beta = \frac{i2}{3} , |\phi_0\rangle_B = \frac{2}{\sqrt{5}}|1\rangle_B + \frac{1}{\sqrt{5}}|0\rangle_B , \text{ and } |\phi_1\rangle_B = |0\rangle_B .$$

The last step required pulling out the factor of $\sqrt{5/3}$ to make sure that $|\phi_0\rangle_B$ is properly normalized for a pure state. Now, that we have rewritten the state, the effect of measuring qubit A can be given as follows:

$$\mathbf{meas}^{(A)}\left(\alpha|\circ\rangle_A|\phi_0\rangle_B + \beta|1\rangle_A|\phi_1\rangle_B\right) = \begin{cases} \circ: |\circ\rangle_A|\phi_0\rangle_B & \text{with probability } |\alpha|^2 \\ 1: |1\rangle_A|\phi_1\rangle_B & \text{with probability } |\beta|^2 \end{cases}. \quad (36)$$

For the example, the measurement outcome is \circ with probability $5/9$, in which case the state collapses to $|\circ\rangle_A(1/\sqrt{5}|\circ\rangle_B + 2/\sqrt{5}|1\rangle_B)$. The outcome is 1 with probability $4/9$, in which case the state collapses to $|1\rangle_A|\circ\rangle_B$. The probabilities add up to 1 as they should.

The same procedure works for figuring out the effect of measuring one of any number of qubits. Say we want to measure qubit B among qubits A, B, C, D, currently in state $|\psi\rangle_{ABCD}$. First, rewrite the state in the form $\alpha|\circ\rangle_B|\phi_0\rangle_{ACD} + \beta|1\rangle_B|\phi_1\rangle_{ACD}$, making sure that the ACD superpositions are pure states. Then, the outcome of the measurement is \circ with probability $|\alpha|^2$ and 1 with probability $|\beta|^2$. The collapsed states are $|\circ\rangle_B|\phi_0\rangle_{ACD}$ and $|1\rangle_B|\phi_1\rangle_{ACD}$, respectively.

Probabilities of the measurement outcomes and the new states can be calculated systematically. For example, to compute the probability and state for outcome \circ of $\mathbf{meas}^{(A)}$ given the state $|\psi\rangle_{AB}$, one can first obtain the unnormalized ket expression $|\phi'_0\rangle_B = {}^A\langle\circ||\psi\rangle_{AB}$ by using the rules for multiplying kets by bras. The probability is given by $p_0 = {}^B\langle\phi'_0|\phi'_0\rangle_B$, and the collapsed, properly normalized pure state is

$$\frac{|\circ\rangle_A|\phi'_0\rangle_B}{\sqrt{p_0}} = \frac{|\circ\rangle_A{}^A\langle\circ||\psi\rangle_{AB}}{\sqrt{p_0}}. \quad (37)$$

The operator $P_\circ = |\circ\rangle_A{}^A\langle\circ|$ is called a projection operator or projector for short. If we perform the same computation for the outcome 1 , we find the projector $P_1 = |1\rangle_A{}^A\langle 1|$. The two operators satisfy $P_a^2 = P_a$, $P_a^\dagger = P_a$, and $P_\circ + P_1 = \mathbb{1}$. In terms of the projectors, the measurement's effect can be written as follows:

$$\mathbf{meas}^{(A)}|\psi\rangle_{AB} = \begin{cases} \circ: P_\circ|\psi\rangle_{AB}/\sqrt{p_0} & \text{with probability } p_0 \\ 1: P_1|\psi\rangle_{AB}/\sqrt{p_1} & \text{with probability } p_1 \end{cases}, \quad (38)$$

where $p_0 = {}^{AB}\langle\psi|P_\circ|\psi\rangle_{AB}$ and $p_1 = {}^{AB}\langle\psi|P_1|\psi\rangle_{AB}$. In quantum mechanics, any pair of projectors satisfying the properties given above is associated with a potential measurement whose effect can be written in the same form. This is called a binary von Neumann, or projective, measurement.

Mixtures and Density Operators

The measurement operation reads out information from qubits to pbits. What if we discard the pbit that contains the measurement outcome? The result is that the qubits are in a probabilistic mixture of two pure states. Such mixtures are a generalization of pure states. The obvious way to think about a mixture is that we have a probability distribution over pure quantum states. For example, after discarding the pbit and qubit A in Equation (36), we can write the state of B as $\rho = \{|\alpha|^2:|\phi_0\rangle_B, |\beta|^2:|\phi_1\rangle_B\}$, using the notation for probability distributions introduced earlier.

Mixtures frequently form when irreversible operations are used, such as measurement. Except for measurement, the quantum gates we have introduced so far are reversible and therefore transform pure states to pure states so that no mixtures can be formed. One of the fundamental results of reversible classical and quantum computation is that there is no loss in power in using only reversible gates. Specifically, it is possible to change a computation that includes irreversible operations to one that accomplishes the same goal, has only reversible operations, and is efficient in the sense that it uses at most polynomial additional resources. However, the cost of using only reversible operations is not negligible. In particular, for ease of programming and, more important, when performing repetitive error-correction tasks (see the article on this subject on page 188), the inability to discard or reset qubits can be very inconvenient. We therefore introduce additional operations that enable resetting and discarding.

Although resetting has a so-called thermodynamic cost (think of the heat generated by a computer), it is actually a simple operation. The **reset** operation applied to qubit A can be thought of as the result of first measuring A, then flipping A if the measurement outcome is $|1\rangle$, and finally discarding the measurement result. Using the notation of Equation (36), the effect on a pure state $|\psi\rangle_{AB}$ is given by

$$\mathbf{reset}^{(A)}|\psi\rangle_{AB} = \left\{ |\alpha|^2 : |0\rangle_A |\phi_0\rangle_B, |\beta|^2 : |0\rangle_A |\phi_1\rangle_B \right\} . \quad (39)$$

To apply **reset** to an arbitrary probability distribution, you apply it to each of that distribution's pure states and combine the results to form an expanded probability distribution. The **discard**^(A) operation is **reset**^(A) followed by discarding qubit A. In the expression for the state after **reset**^(A), therefore, all the $|0\rangle_A$ are removed. It is an important fact that every physically realizable quantum operation, whether reversible or not, can be expressed as a combination of **add** operations, gates from the universal set, and **discard** operations.

The representation of mixtures using probability distributions over pure states is redundant. That is, many probability distributions are physically indistinguishable. A nonredundant description of a quantum state can be obtained if density operators are used. The density operator for the mixture ρ in Equation (39) is given by

$$\hat{\rho} = |\alpha|^2 |\phi_0\rangle_B \langle\phi_0| + |\beta|^2 |\phi_1\rangle_B \langle\phi_1| . \quad (40)$$

The general rule for calculating the density operator from a probability distribution is the following: For each pure state $|\phi\rangle$ in the distribution, calculate the operators $|\phi\rangle\langle\phi|$ and sum them weighted by their probabilities.

There is a way to apply gates to the density operators defined by states. If the gate acts by the unitary operator U , then the effect of applying it to $\hat{\rho}$ is given by $U\hat{\rho}U^\dagger$ where U^\dagger is the conjugate transpose of U . (In the bra-ket expression for U , U^\dagger is obtained by replacing all complex numbers by their conjugates, and terms such as $|\phi\rangle\langle\phi|$, by $\langle\phi|\langle\phi|$.)

The relationship between a qubit's state space and a sphere can be explained in terms of the qubit's density operators. In matrix form, this operator is a 2×2 matrix, which can be written uniquely as a sum $(\mathbb{1} + x\sigma_x + y\sigma_y + z\sigma_z)/2$. One can check that, if the density operator $|\psi\rangle\langle\psi|$ for a qubit's pure state is written as such a sum,

$$|\psi\rangle\langle\psi| = (\mathbb{1} + x\sigma_x + y\sigma_y + z\sigma_z)/2 \quad , \quad (41)$$

then the vector (x, y, z) thus obtained is on the surface of the unit sphere in three dimensions. In fact, for every vector (x, y, z) on the unit sphere, there is a unique pure state satisfying Equation (41). Since the density operators for mixtures are arbitrary, convex (that is, probabilistic) sums of pure states, the set of (x, y, z) thus obtained for mixtures fills out the unit ball. The rotations introduced earlier modify the vector (x, y, z) in the expected way, by rotation of the vector around the appropriate axis. See the article "NMR and Quantum Information Processing" on page 232 for more details.

Quantum Computation

The model of computation defined by the one- and two-qubit gates and the operations **add**, **meas**, and **discard** qubits is called the quantum network model. A sequence of instructions for applying these operations is called a quantum network. Quantum computation extends the network model by providing a means for repeating blocks of instructions. Such means can be specified by a formal machine model of computation. There are several such models of classical and quantum computers. One of the best known is the Turing machine, which has a quantum analogue, the quantum Turing machine. This model has its uses for formal studies of computation and complexity but is difficult to program. Fortunately, as mentioned in Part I, there is no loss of computational power if the means for repeating instructions is provided by a classical computer that can apply gates and other operations to qubits. A general quantum algorithm is a program written for such a computer.

There are three practical methods that can be used to write quantum networks and algorithms. The first is to use the names for the different operations and algebraically multiply them. The second is to draw quantum networks, which are pictorial representations of the sequence of steps in time, somewhat like flowcharts without loops. The third is to use a generic programming language enhanced with statements for accessing and modifying quantum bits. The first two methods work well as long as the sequence is short and we do not use many operations that depend on measurement outcomes or require loops. They are often used to describe subroutines of longer algorithms presented either in words or by use of the third method.

To see how to use the different methods and also to illustrate the power of quantum computation, we work out a short quantum algorithm that solves the so-called parity problem.

The Parity Problem. Given is a “black-box” quantum operation $\mathbf{BB}^{(ABC)}$ that has the following effect when applied to a logical basis state:

$$\mathbf{BB}^{(ABC)}|a_A a_B\rangle_{AB}|a_C\rangle_C = |a_A a_B\rangle_{AB}|a_C \oplus (b_A a_A \oplus b_B a_B)\rangle_C, \quad (42)$$

where b_A and b_B are 0 or 1. The actual values of b_A and b_B are unknown. The problem is to determine what b_A and b_B are by using the black box only once.

The terminology and definition of the operation $\mathbf{BB}^{(ABC)}$ require explanation. In computation, we say that an operation is a black box, or an “oracle,” if we have no access whatsoever to how the operation is implemented. In a black-box problem, we are promised that the black box implements an operation from a specified set of operations. In the case of the parity problem, we know that the operation is to add one of four possible parities (see below). The problem is to determine that parity by using the black box in a network. Black-box problems serve many purposes. One is to study the differences between models of computation, just as we are about to do. In fact, black-box problems played a crucial role in the development of quantum algorithms by providing the first and most convincing examples of the power of quantum computers (Bernstein and Vazirani 1993, Simon 1994). Some of these examples involve generalizations of the parity problem. Another purpose of black-box problems is to enable us to focus on what can be learned from the input/output behavior of an operation without having to analyze its implementation. Focusing on the input/output behavior is useful because, in many cases of interest, it is difficult to exploit knowledge of the implementation in order to determine a desirable property of the operation. A classical example is the well-known satisfiability problem, in which we are given a classical circuit with one output bit and we need to determine whether there is an input for which the output is 1. Instead of trying to analyze the circuit, one can look for and use a general-purpose black-box search algorithm to find the satisfying input.

In the definition of the effect of $\mathbf{BB}^{(ABC)}$, the operation \oplus is addition modulo 2, so $1 \oplus 1 = 0$, and all the other sums are as expected. As the state symbols have a numeric meaning now, we will use the number font for states. To see what \mathbf{BB} does, suppose that b_A and b_B are both 1. Then \mathbf{BB} adds (modulo 2) the parity of the logical state in AB to the logical state of C. The parity of a logical state is 0 if the number of 1s is even and 1 if it is odd. The action of \mathbf{BB} for this example is given by

$$\begin{aligned} \mathbf{BB}^{(ABC)}|00\rangle_{AB}|0\rangle_C &= |00\rangle_{AB}|0\rangle_C \\ \mathbf{BB}^{(ABC)}|01\rangle_{AB}|0\rangle_C &= |01\rangle_{AB}|0 \oplus 1\rangle_C \\ &= |01\rangle_{AB}|1\rangle_C \\ \mathbf{BB}^{(ABC)}|10\rangle_{AB}|1\rangle_C &= |10\rangle_{AB}|1 \oplus 1\rangle_C \\ &= |10\rangle_{AB}|0\rangle_C \\ \mathbf{BB}^{(ABC)}|11\rangle_{AB}|0\rangle_C &= |11\rangle_{AB}|0\rangle_C \end{aligned} \quad (43)$$

The action of the black box is extended to superpositions by linear extension. This means that to apply **BB** to a superposition of the logical states, we simply apply it to each logical summand and add the results. Different values of b_A and b_B correspond to different parities. For example, if $b_A = 1$ and $b_B = 0$, then the parity of the state in A is added to the state in C. In this sense, what is added is the parity of a subset of the two qubits AB. Thus, one way of thinking about the problem is that we wish to find out which subset's parity the black box is using.

We can give an algorithm that solves the parity problem using each of the three methods for describing quantum networks. Here is an algebraic description of a solution, **qparity**^(ABC), given as a product of quantum gates that involves one use of the black box. We defer the explanation of why this solution works until after we show how to represent the algorithm pictorially, using quantum networks.

$$\mathbf{qparity}^{(ABC)} = \mathbf{meas}^{(B)} \mathbf{H}^{(B)} \mathbf{meas}^{(A)} \mathbf{H}^{(A)} \mathbf{BB}^{(ABC)} \mathbf{H}^{(C)} \mathbf{not}^{(C)} \mathbf{add}^{(C)} \mathbf{H}^{(B)} \mathbf{add}^{(B)} \mathbf{H}^{(A)} \mathbf{add}^{(A)} . \quad (44)$$

The output of the algorithm is given by the classical outputs of the measurements of qubit A, which yield b_A , and of qubit B, which yield b_B . As is conventional, in writing products of linear operators, the order of application in Equation (44) is right to left, as in a product of matrices applied to a column vector. This order of terms in a product is, however, counterintuitive, particularly for operations to be performed sequentially. It is therefore convenient to use left to right notation, as is done in describing laser or radio-frequency pulse sequences, and to put dots between gates to indicate left to right order:

$$\mathbf{qparity}^{(ABC)} = \mathbf{add}^{(A)} \mathbf{.H}^{(A)} \mathbf{.add}^{(B)} \mathbf{.H}^{(B)} \mathbf{.add}^{(C)} \mathbf{.not}^{(C)} \mathbf{.H}^{(C)} \mathbf{.BB}^{(ABC)} \mathbf{.H}^{(A)} \mathbf{.meas}^{(A)} \mathbf{.H}^{(B)} \mathbf{.meas}^{(B)} . \quad (45)$$

In this representation, the first operation is **add**^(A), the second is **H**^(A) (the Hadamard gate on qubit A), and so on.

The algebraic specification of the algorithm as products of gates does not make it easy to see why the algorithm works. It is also difficult to see which operations depend on each other. Such dependencies are used to determine whether the operations can be parallelized. Quantum networks make these tasks simpler. The quantum network for the above sequence is shown in Figure 2.

To understand how the quantum network illustrated in Figure 2 solves the parity problem, we can follow the states as the network is executed from left to right, using the indicated checkpoints. Using vector notation for the states, at checkpoint 1, the state is

$$|\psi\rangle_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (46)$$

where we used Kronecker product notation to denote the states of A, B, and C in this order. In the next time step, the network involves applying Hadamard gates—see Equation (13)—to A and B and a **not** gate—see Equation (9)—to C. At checkpoint 2, this operation results in the state

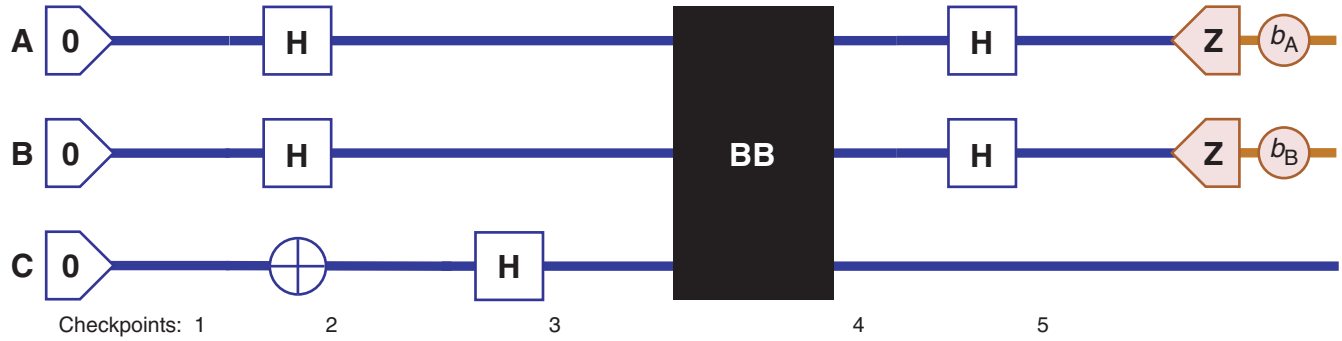


Figure 2. Quantum Network for Solving the Parity Problem

A quantum network has a line (horizontal in this case) for each qubit. The line can be thought of as the timeline for the qubit and is shown in blue. Each gate is drawn as a box, circle, or other element intercepting the lines of the qubits it acts on. In this case, time runs from left to right. Each qubit’s timeline starts at the point where it is added. In this example, the qubits’ timelines end when they are measured, at which point a classical bit (brown timeline) containing the measurement outcome is introduced. The operation BB is illustrated as a black box. The numbers underneath the network refer to checkpoints used to explain how the network solves the parity problem.

$$|\psi\rangle_2 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \otimes \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{47}$$

Next, a Hadamard gate is applied to C, so that at checkpoint 3, we have

$$|\psi\rangle_3 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \otimes \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \otimes \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}. \tag{48}$$

The next event involves applying the black box. To understand what happens, note that the effect of the black box can be described as, “apply **not** to C if the parity according to b_A and b_B of the logical state of AB is 1.” The current state of C is such that, if **not** is applied, only the sign changes:

$$\begin{aligned} \text{not} \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \\ &= - \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}. \end{aligned} \tag{49}$$

Now, AB is in a superposition of each of the logical states, and conditional on the logical state and the (hidden) parity, the sign changes. As a result, although the state of C does not change, a phase is “kicked back” to AB. A generalization of this effect is at the heart of Alexei Kitaev’s version of Peter Shor’s quantum factoring algorithm (see the article “From Factoring to Phase Estimation” on page 38). At the next checkpoint, and after some arithmetic to check which logical states change sign, we can write the state as

$$|\psi\rangle_4 = \begin{pmatrix} 1/\sqrt{2} \\ (-1)^{b_A}/\sqrt{2} \end{pmatrix} \otimes \begin{pmatrix} 1/\sqrt{2} \\ (-1)^{b_B}/\sqrt{2} \end{pmatrix} \otimes \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}. \tag{50}$$

Notice that qubits A and B are in orthogonal states for different values of b_A and b_B . It suffices to apply the Hadamard transform again to A and B to get

$$|\psi\rangle_4 = \begin{pmatrix} 1-b_A \\ b_A \end{pmatrix} \otimes \begin{pmatrix} 1-b_B \\ b_B \end{pmatrix} \otimes \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}. \quad (51)$$

Measurements of A and B now reveal the previously unknown b_A and b_B .

As can be seen, the visual representation of a quantum network eases the tasks of following what happens. This is why it is used extensively for presenting basic subroutines and algorithms in quantum computation. A guide to the commonly used network elements is given in Table I.

When designing or describing complicated algorithms for quantum computers, providing everything in terms of quantum networks can become difficult, particularly when an important part of the algorithm consists of computations that are best done on a classical computer. For example, a full description of Shor's algorithm for factoring integers (see the article "From Factoring to Phase Estimation" on page 38) includes a significant amount of classical preprocessing, which determines choices made in the quantum algorithm, and classical postprocessing, which computes a factor from the measured result by a continued fraction algorithm. For such algorithms, one can use a programming language similar to Pascal, BASIC, or C enhanced with statements to access quantum bits and to apply quantum operations. For algorithm design, computer scientists often use a semiformal language called pseudocode (Cormen et al. 1990). With a simple extension called quantum pseudocode, the algorithm for the parity problem can be written as follows:

BBPARITY(BB)

Input: Access to a quantum black box **BB** that acts on three qubits by adding a parity function of the first two qubits to the third

Output: The two bits b_A and b_B of the parity function

```

foreach  $i \in \{A, B, C\}$ 
   $\lceil a_i \rceil \leftarrow |0\rangle$ 
  C: Initialize three one-qubit registers  $\lceil a_i \rceil$ ,  $i = A, B, C$ .

  The corner bracket annotation declares  $a_i$  as a quantum register.

  end
   $\lceil a_C \rceil \leftarrow \sigma_x \lceil a_C \rceil$ 
  foreach  $i \in \{A, B, C\}$ 
     $\lceil a_i \rceil \leftarrow \mathbf{H} \lceil a_i \rceil$ 
  end
   $\lceil a \rceil \leftarrow \mathbf{BB} \lceil a \rceil$ 
  C:  $\lceil a \rceil$  refers to the three-qubit register consisting of the  $\lceil a_i \rceil$ .
  foreach  $i \in \{A, B\}$ 
     $\lceil a_i \rceil \leftarrow \mathbf{H} \lceil a_i \rceil$ 

```

```

 $b_i \leftarrow \text{meas } \lceil a_i \rceil$ 
end

return  $b_A, b_B$ 

end
    
```

Any classical programming language can be extended with statements to access and manipulate quantum registers.

Now, that we have looked at the quantum solution to the parity problem, let us consider the question of the least number of black-box applications required by a classical algorithm: Each classical use of the black box can only give us one bit of information. In particular, one use of the black box with input $a_A a_B$ reveals only the parity of $a_A a_B$ according to the hidden parameters b_A and b_B . Each use of the black box can therefore only help us distinguish between two subsets of the four possible parities. At least two uses of the black box are therefore necessary. Two uses are also sufficient. To determine which of the four parities is involved, use the black box first with input $a_A a_B = 10$ and then with input $a_A a_B = 01$. As a result of this argument, one can consider the parity problem as a simple example of a case in which there is a more efficient quantum algorithm than is possible

Table I. Quantum Network Elements



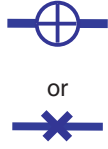





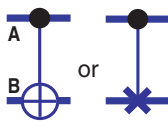
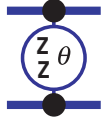
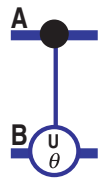
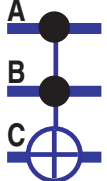
Gate Names and Their Abbreviations	Gate Symbols	Algebraic Form	Matrix Form
Add/Prepare, add		If applied to an existing qubit $\{ 0\rangle\langle 0 , 0\rangle\langle 1 \}$ (operator mixture)	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ or $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$
Measure, meas		$\{0: 0\rangle\langle 0 , 1: 1\rangle\langle 1 \}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$
Not, not , σ_x		$ 0\rangle\langle 1 + 1\rangle\langle 0 $	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
Hadamard, H		$e^{-i\sigma_y\pi/4} \sigma_z$	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$
Phase Change, S ($e^{i\phi}$)		$e^{i\phi/2} e^{-i\sigma_z\phi/2}$	$\begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}$

Table I. (Continued)

Gate Names and Their Abbreviations	Gate Symbols	Algebraic Form	Matrix Form
z-Rotation, Z_ϕ		$e^{-i\sigma_z\phi/2}$	$\begin{pmatrix} e^{-i\phi/2} & 0 \\ 0 & e^{i\phi/2} \end{pmatrix}$
y-Rotation, Y_θ		$e^{-i\sigma_y\theta/2}$	$\begin{pmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{pmatrix}$
x-Rotation, X_θ		$e^{-i\sigma_x\theta/2}$	$\begin{pmatrix} \cos(\theta/2) & -i\sin(\theta/2) \\ -i\sin(\theta/2) & \cos(\theta/2) \end{pmatrix}$
Controlled not, cnot		$ 0\rangle_A \langle 0 + 1\rangle_A \langle 1 \sigma_x^{(B)}$ $e^{-i\sigma_z^{(A)}\pi/4} e^{-i/2(1-\sigma_z^{(A)})\sigma_x^{(B)}\pi/2}$	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$
zz-Rotation, (ZZ) $_\theta$		$e^{-i\sigma_z^{(A)}\sigma_z^{(B)}\theta/2}$	$\begin{pmatrix} e^{-i\theta/2} & 0 & 0 & 0 \\ 0 & e^{i\theta/2} & 0 & 0 \\ 0 & 0 & e^{i\theta/2} & 0 \\ 0 & 0 & 0 & e^{-i\theta/2} \end{pmatrix}$
Controlled Rotation, cU $_\theta$		$ 0\rangle_A \langle 0 + 1\rangle_A \langle 1 e^{-i\sigma_U^{(B)}\theta/2}$	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & & e^{-i\sigma_U\theta/2} \\ 0 & 0 & & \end{pmatrix}$
Toffoli Gate, c²not		$\mathbb{1} - 11\rangle_{AB} \langle 11 + 11\rangle_{AB} \langle 11 \sigma_x^{(C)}$	

classically. However, it is worth noting that the comparison is not entirely fair: A truly classical oracle answering parity questions or implementing the black box on the states of classical bits is useless to a quantum algorithm. To take advantage of such an algorithm, it must be possible to use superpositions that are not implicitly collapsed. Collapse can happen if the oracle makes a measurement or otherwise “remembers” the question that it was asked.

Resource Accounting

When trying to solve a problem using quantum information processing, an important issue is to determine what physical resources are available and how much of each resource is needed for the solution. As mentioned before, in classical information, the primary resources are bits and operations. The number of bits used by an algorithm is its space requirement; the number of operations used, its time requirement. If parallel computation is available, one can distinguish between the total number of operations (work) and the number of parallel steps (time).

When quantum information processing is used, the classical resources are still relevant for running the computer that controls the quantum system and performs any preprocessing and postprocessing tasks. The main quantum resources are analogous to the classical ones: Quantum space is the number of qubits needed, and quantum time, the number of quantum gates. Because it turns out that reset operations have a thermodynamic cost, one can count irreversible quantum operations separately. This accounting of the resource requirements of algorithms and of the minimum resources needed to solve problems forms the foundation of quantum complexity theory.

As a simple example of resource accounting, consider the algorithm for the parity problem. No classical computation is required to decide which quantum gates to apply or to determine the answer from the measurement. The quantum network consists of a total of 11 quantum gates (including **add** and **meas** operations) and one oracle call (the application of the black box). In the case of oracle problems, one usually counts the number of oracle calls first, as we have done in discussing the algorithm. The network is readily parallelized to reduce the time resource to 6 steps.

Part III: Advantages of Quantum Information

The notion of quantum information as explained in this primer was established in the 1990s. It emerged from research focused on understanding how physics affects our capabilities to communicate and process information. The recognition that usable types of information need to be physically realizable was repeatedly emphasized by Rolf Landauer, who proclaimed that “information is physical” (1991). Beginning in the 1960s, Landauer studied the thermodynamic cost of irreversible operations in computation (1961). Charles Bennett showed that, by using reversible computation, this cost can be avoided (1973). Limitations of measurement in quantum mechanics were investigated early by researchers such as John von Neumann (1932a and 1932b) and later by Alexander Holevo (1973b) and Carl Helstrom (1976). Holevo introduced the idea of quantum communication channels and found bounds on their capacity for transmitting classical information (1973a). Initially, most work focused on determining the physical limitations placed on classical information processing. The fact that pairs of two-level systems can have correlations not possible for classical systems was proved by John Bell (1964). Subsequently, indications that quantum mechanics offers advantages to information processing came from Stephen Wiesner’s studies of cryptographic applications in the late 1960s. Wiesner’s work was not recognized, however, until the 1980s,

when Bennett, Gilles Brassard, Seth Breidbart, and Wiesner introduced (1982) the idea of quantum cryptography, which can be used to communicate in secret.

Initially, the term quantum computation was mostly used to refer to classical computers realized with quantum mechanical systems. In the 1980s, Paul Benioff (1980), Richard Feynman (1982), and Yuri Manin (1980) introduced the idea of a quantum computer based on quantum information. They noted that the apparent exponential complexity of simulating quantum mechanics on a classical computer might be overcome if one could use a computer based on quantum mechanics. A formal model of quantum Turing machines was soon defined by David Deutsch (1985), who later also introduced quantum networks (1989). Deutsch and Richard Jozsa (1992) were the first to introduce a black-box problem that could be solved deterministically on a quantum computer in fewer steps than on a classical computer.

In spite of suggestions that it could lead to large efficiency improvements in simulating physics, quantum information processing was still largely an academic subject. Based on work by Ethan Bernstein and Umesh Vazirani (1993) that formalized quantum complexity theory, Dan Simon (1994) showed that, for black-box problems, quantum computers can be exponentially more efficient than classical deterministic or probabilistic computers, giving the first indication of a strong advantage for quantum information processing. It was Shor's algorithm for factoring large integers (1994 and 1997) that finally convinced a larger community that quantum information was more than just a tool for realizing classical computers. This change in attitude was in no small part due to the fact that the security of commonly used cryptographic protocols is based on the difficulty of factoring.

At that point, it was still generally believed that the fragility of quantum states made it unlikely for reasonably large quantum computers to be realized in practice. But the discovery by Shor (1995) and Andrew Steane (1996) that quantum error correction was possible soon changed that view (for an introductory overview, see the article on quantum error correction on page 188).

Because the usefulness and realizability of quantum information has been recognized, the science of quantum information processing is a rapidly growing field. As quantum information becomes increasingly accessible by technology, its usefulness will be more apparent. The next few sections discuss what we currently know about applications of quantum information processing. Refer to Michael Nielsen and Isaac Chuang (2001) as a useful reference text on quantum computation and information with historical notes.

Quantum Algorithms

Shor's factoring algorithm, which precipitated much of the current work in quantum information processing, is based on a quantum realization of the fast Fourier transform. The most powerful version of this technique is now represented by the phase estimation algorithm of Kitaev (1995) as formalized by Richard Cleve et al. (1998). (For an explanation, see the article "From Factoring to Phase Estimation" on page 38.) The best-known application of quantum factoring is cryptanalysis, where it allows efficiently breaking the currently used public-key cryptographic codes. Whether there are any constructive applications of quantum factoring and its generalizations remains to be determined. For users of public-key cryptography, a crucial question is, "How long can public-key codes based on factoring continue to be used safely?" To attempt an answer to this question, one can note that to break a code with a typical key size of 1000 bits requires more than 3000 qubits and 10^8 quantum gates, which is well out of reach of current technology. However, it is conceivable that a recording of encrypted information transmitted in 2000 can be broken in the next "few" decades.

Shor's quantum factoring algorithm was not the first with a significant advantage over classical algorithms. The first proposed quantum algorithms with this property were for simulating quantum mechanical systems. These algorithms simulate the evolution of a reasonably large number of interacting quantum particles—for example, the electrons and nuclei in a molecule. The algorithms' outputs are what would be measurable physical quantities of the system being simulated. The known methods for obtaining these quantities on classical computers scale exponentially with the number of particles, except in special cases.

The idea of using quantum computers for simulating quantum physics spurred the work that eventually led to the quantum factoring algorithm. However, that idea did not have the broad scientific impact that the quantum factoring algorithm had. One reason is that, because of its cryptographic applications, factoring is a heavily studied problem in theoretical computer science and cryptography. Because so many people have tried to design efficient algorithms for factoring and failed, the general belief that factoring is hard for classical computers has a lot of credibility. In contrast, a quantum physics simulation has no simple formulation as an algorithmic problem suitable for study in theoretical computer science. Furthermore, many researchers still believe that the physically relevant questions can be answered with efficient classical algorithms, requiring only more cleverness on the part of algorithm designers. Another reason for the lack of impact is that many of the fundamental physical quantities of interest are not known to be efficiently accessible even on quantum computers. For example, one of the first questions about a physical system with a given Hamiltonian (energy observable) is, "What is the ground-state energy?" It is known that the ability to efficiently answer this question for physically reasonable Hamiltonians leads to efficient algorithms for hard problems, such as the traveling salesman or the scheduling problems. In spite of occasional claims to the contrary, an efficient quantum solution to these problems is widely considered unlikely.

Most quantum algorithms for physics simulations are based on a direct emulation of a quantum mechanical system's evolution. The focus of the original proposals by Feynman and others was on how to implement the emulation using a suitable formulation of general-purpose quantum computers. After such computers were formalized by Deutsch, the implementation of the emulation was generalized and refined by Seth Lloyd (1996), Wiesner (1996), and Christof Zalka (1998). The ability to emulate the evolution of quantum systems is actually widely used by classical Monte Carlo algorithms for simulating physics. In those algorithms, state amplitudes are, in effect, represented by expectations of random variables that are computed during the simulation. As in the case of quantum algorithms for physics emulation, Monte Carlo algorithms efficiently evolve the representation of the quantum system. The inefficiency of the classical algorithm arises only in determining a physical quantity of interest. In the case of Monte Carlo algorithms, the measurement of a physical quantity suffers from the so-called sign problem, often resulting in exponentially large, random errors that can be reduced only by repeating the computation exponentially many times. In contrast, the quantum algorithms for emulation can determine many (but not all) of the interesting physical quantities with polynomially bounded statistical errors. How to efficiently implement measurements of these quantities has been the topic of more recent work in this area, much of which is based on variants of the phase-estimation algorithm (Terhal and DiVincenzo 2000, Knill and Laflamme 1998, Abrams and Lloyd 1999, Ortiz et al. 2001, Miquel et al. 2002).

Although several researchers have suggested that there are interesting quantum physics simulations that can be implemented with well below 100 qubits, one of the interesting problems in this area of research is to come up with a specific simulation algorithm using small numbers of qubits and quantum gates, an algorithm that computes an interesting physical quantity not easily obtainable using available classical computers.

Another notable algorithm for quantum computers, unstructured quantum search, was described by Lov Grover (1996). Given is a black box that computes a binary function f on inputs x with $0 \leq x < N$. The function f has the property that there is a unique input a for which $f(a) = 1$. The standard quantum version of this black box implements the transformation $\hat{f}|x\rangle|b\rangle = |x\rangle|b \oplus f(x)\rangle$, where b is a bit and $b \oplus f(x)$ is computed modulo 2. Unstructured quantum search finds a quadratically faster, that is, in time of order $N^{1/2}$, than the best classical black-box search, which requires time of order N . The context for this algorithm is the famous $P \neq NP$ conjecture, which is captured by the following algorithmic problem: Given is a classical circuit C that computes an output. Is there an input to the circuit for which the circuit's output is 1? Such an input is called a satisfying input or assignment. For any given input, it is easy to check the output, but an efficient algorithm that finds a satisfying input is conjectured to be impossible. This is the $P \neq NP$ conjecture. Generalizations of Grover's search algorithm—amplitude amplification (Brassard et al. 1998)—allow finding satisfying inputs faster than naive, classical search does, which tries every possible input in some, possibly random, order. It is worth noting, however, that if sufficient classical parallelism is available, quantum search loses many of its advantages.

The three algorithms just described capture essentially all the known algorithmic advantages of quantum computers. Almost all algorithms that have been described are applications of phase estimation or of amplitude amplification. These algorithms well justify developing special-purpose quantum information-processing technology. Will general-purpose quantum computers be useful? More specifically, what other algorithmic advantages do quantum computers have?

Quantum Communication

Quantum communication is an area in which quantum information has proven (rather than conjectured) advantages. The best-known application is quantum cryptography, whereby two parties, Alice and Bob, can generate a secret key using a quantum communication channel (for example, photons transmitted in optical fiber) and an authenticated classical channel (for example, a telephone line). Any attempt at learning the key by eavesdropping is detected. A quantum protocol for generating a secret key is called a quantum-key-exchange protocol. There are no equally secure means for generating a secret key by using only classical deterministic channels. Few quantum operations are needed to implement quantum key exchange, and as a result, there are working prototype systems (Hughes et al. 2000, Townsend 1998, Ribordy et al. 2001). To overcome the distance limitations (tens of kilometers) of current technology requires the use of quantum error correction and hence more demanding quantum technology.

Quantum key exchange is one of an increasing number of multiparty problems that can be solved more efficiently with quantum information. The area of research concerned with how several parties at different locations can solve problems while minimizing communication resources is called communication complexity. For quantum communication complexity (Cleve and Burhman 1997), the communication resources include either shared entangled qubits or a means for transmitting quantum bits. A seminal paper by Howard Burhman, Cleve, and Wim van Dam (2000) shows how the nonclassical correlations present in maximally entangled states lead to protocols based on such states that are more efficient than any classical deterministic or probabilistic protocol achieving the same goal. Ran Raz (1999) showed that there is an exponential improvement in communication resources for a problem in which Alice and Bob have to answer a question about the relationship between a vector known to Alice and a matrix known to Bob. Although this problem is artificial, it suggests that there are potentially useful advantages to be gained from quantum information in this setting.

Quantum Control

According to Moore's law of semiconductor technology, the size of transistors is decreasing exponentially, by a factor of about .8 every year. If this trend continues, then over the next few decades, devices will inevitably be built whose behavior will be primarily quantum mechanical. For the purpose of classical computation, the goal is to remove the quantum behavior and stabilize classical information. But quantum information offers an alternative: It is possible to directly use quantum effects to advantage. Whether or not this alternative is useful (and we believe it is), the ideas of quantum information can be used to systematically understand and control quantum mechanical systems.

The decreasing size of semiconductor components is a strong motivation to strive for better understanding the behavior of condensed-matter quantum mechanical systems. But there is no reason to wait for Moore's law: There are a rapidly increasing number of experimental systems in which quantum mechanical effects are being used and investigated. Examples include many optical devices (lasers, microwave cavities, entangled photon pairs), nuclear magnetic resonance with molecules or in solid state, trapped ion or atom systems, Rydberg atoms, superconducting devices (Josephson junctions and SQUIDs), and spintronics (electron spins in semiconductor devices). Many of these systems are being considered as candidates for realizing quantum information processing. Yet, regardless of the future of quantum information processing, there is ample motivation for studying these systems.

Outlook

The science of quantum information processing is promising a significant impact on how we process information, solve algorithmic problems, engineer nanoscale devices, and model fundamental physics. It is already changing the way we understand and control matter at the atomic scale, making the quantum world more familiar, accessible, and understandable. Whether or not we do most of our everyday computations by using the classical model, it is likely that the physical devices that support these computations will exploit quantum mechanics and integrate the ideas and tools that have been developed for quantum information processing. ■

Acknowledgment

We thank Nikki Cooper and Ileana Buican for their extensive encouragement and editorial help.

Further Reading

- Abrams, D. S., and S. Lloyd. 1999. Quantum Algorithm Providing an Exponential Speed Increase for Finding Eigenvalues and Eigenvectors. *Phys. Rev. Lett.* **83**: 5162.
- Barenco, A., C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, et al. 1995. Elementary Gates for Quantum Computation. *Phys. Rev. A* **52**: 3457.
- Bell, J. S. 1964. On the Einstein-Podolsky-Rosen Paradox. *Phys.* **1**: 195.
- Benioff, P. 1980. The Computer as a Physical System: A Microscopic Quantum Mechanical Hamiltonian Model of Computers as Represented by Turing Machines. *J. Stat. Phys.* **22**: 563.
- Bennett, C. H. 1973. Logical Reversibility of Computation. *IBM J. Res. Dev.* **17**: 525.
- Bennett, C. H., G. Brassard, S. Breidbart, and S. Wiesner. 1982. Quantum Cryptography, or Unforgeable Subway Tokens. In *Advances in Cryptology: Proceedings of Crypto '82*. Edited by D. Chaun, R. L. Rivest, and A. T. Sherman, p. 267. New York: Plenum Press.



Emanuel (Manny) Knill received his Ph. D. in pure mathematics from the University of Colorado at Boulder in 1991. Since 1992, he has been with Los Alamos National Laboratory. Manny has worked on various aspects of quantum information processing since 1995. (Photo of Augustus)

Contact Information

E. Knill: knill@lanl.gov

R. Laflamme: laflamme@iqc.ca

H. Barnum: barnum@lanl.gov

D. Dalvit: dalvit@lanl.gov

J. Dziarmaga: ufjacekd@th.if.uj.edu.pl

J. Gubernatis: jg@lanl.gov

L. Gurvits: gurvits@lanl.gov

G. Ortiz: g_ortiz@lanl.gov

L. Viola: lviola@lanl.gov

W. Zurek: whz@lanl.gov

- Bernstein, E., and U. Vazirani. 1993. Quantum Complexity Theory. In *Proceedings of the 25th Annual ACM Symposium on the Theory of Computing 1993*, p. 11. New York: ACM Press.
- Bolker, E. D. 1970. *Elementary Number Theory: An Algebraic Approach*. New York: W. A. Benjamin, Inc.
- Brassard, G., P. Hoyer, and A. Tapp. 1998. Quantum Counting. In *Automata, Languages and Programming, Proceedings of ICALP '98*, Vol. 1443 of *Lecture Notes in Computer Science*. Edited by K. G. Larsen, S. Skyum, and G. Winskel, p. 820. Berlin: Springer Verlag.
- Burhman, H., R. Cleve, and W. Van Dam. 2000. Quantum Entanglement and Communication Complexity. *SIAM J. Comput.* **30**: 1829.
- Cleve, R., and H. Buhrman. 1997. Substituting Quantum Entanglement for Communication. *Phys. Rev. A* **56**: 1201.
- Cleve, R., A. Ekert, C. Macchiavello, and M. Mosca. 1998. Quantum Algorithms Revisited. *Proc. R. Soc. London, Ser. A* **454**: 339.
- Cormen, T. H., C. B. Leiserson, and R. L. Rivest. 1990. *Introduction to Algorithms*. Cambridge, MA: MIT Press.
- Deutsch, D. 1985. Quantum Theory, The Church-Turing Principle and the Universal Quantum Computer. *Proc. R. Soc. London, Ser. A* **400**: 97.
- . 1989. Quantum Computational Networks. *Proc. R. Soc. London, Ser. A* **425**: 73.
- Deutsch, D., and R. Jozsa. 1992. Rapid Solution of Problems by Quantum Computation. *Proc. R. Soc. London, Ser. A* **439**: 553.
- Ekert, A. 1998. From Quantum Code-Making to Quantum Code-Breaking. In *The Geometric Universe*, p. 195. Oxford: Oxford University Press.
- Feynman, R. P. 1982. Simulating Physics with Computers. *Int. J. Theor. Phys.* **21**: 467.
- Griffiths, R. B., and C.-S. Niu. 1996. Semiclassical Fourier Transform for Quantum Computation. *Phys. Rev. Lett.* **76**: 3228.
- Grover, L. K. 1996. A Fast Quantum Mechanical Algorithm for Database Search. In *Proceedings of the 28th Annual ACM Symposium on the Theory of Computation*, p. 212, New York: ACM Press.
- Gupta, R., S. A. Smolka, and S. Bhaskar. 1994. On Randomization in Sequential and Distributed Algorithms. *ACM Comput. Surveys* **26**: 7.
- Hardy, G. H., and E. M. Wright. 1979. *An Introduction to the Theory of Numbers*. Fifth edition. London: Oxford University Press.
- Helstrom, C. W. 1976. *Quantum Detection and Estimation Theory. Mathematics in Science and Engineering* Vol. 123. New York: Academic Press.
- Holevo, A. S. 1973a. Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel. *Probl. Inf. Transm.* **9**: 177.
- . 1973b. Statistical Problems in Quantum Physics. In *Proceedings of the Second Japan-USSR Symposium on Probability Theory, Lecture Notes in Mathematics* Vol. 330. Edited by G. Maruyama and Y. V. Prokhorov, p. 104. Berlin: Springer Verlag.
- Hughes, R. J., G. L. Morgan, and C. G. Peterson. 2000. Quantum Key Distribution Over a 48-km Optical Fibre Network. *J. Modern Opt.* **47**: 533.
- Kitaev, A. Yu. 1995. Quantum Measurements and the Abelian Stabilizer Problem. [Online]: [http://eprints.lanl.gov. \(quant-ph/9511026\)](http://eprints.lanl.gov. (quant-ph/9511026)).
- Knill, E., and R. Laflamme. 1998. On the Power of One Bit of Quantum Information. *Phys. Rev. Lett.* **81**: 5672.
- Landauer, R. 1961. Irreversibility and Heat Generation in the Computing Process. *IBM J. Res. Dev.* **5**: 183.
- . 1991. Information is Physical. *Phys. Today* **44**: 22.
- Lloyd, S. 1996. Universal Quantum Simulators. *Science* **273**: 1073.
- Manin, Y. I. 1980. *The Computable and the Not Computable*. Moscow: Sovetskoye Radio. (In Russian).
- Miquel, C., J. P. Paz, M. Saraceno, E. Knill, R. Laflamme, and C. Negrevergne. 2002. Interpretation of Tomography and Spectroscopy as Dual Forms of Quantum Computations. *Nature* **418**: 59.
- Nielsen, M. A., and I. L. Chuang. 2001. *Quantum Computation and Quantum Information*. Cambridge: Cambridge University Press.

- Ortiz, O., J. E. Gubernatis, E. Knill, and R. Laflamme. 2001. Quantum Algorithms for Fermionic Simulations. *Phys. Rev. A* **64**: 022319.
- Papadimitriou, C. H. 1994. *Computational Complexity*. Reading, MA: Addison-Wesley.
- Raz, R. 1999. Exponential Separation of Quantum and Classical Communication Complexity. In *Proceedings of the 31st Annual ACM Symposium on the Theory of Computation (STOC)*, p. 358. El Paso, TX: ACM Press.
- Ribordy, O., J. Brendel, J.-D. Gautier, N. Gisin, and H. Zbinden. 2001. Long-Distance Entanglement-Based Quantum Key Distribution. *Phys. Rev. A* **63**: 012309.
- Shor, P. W. 1994. Algorithms for Quantum Computation: Discrete Logarithms and Factoring. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*. p. 124. Los Alamitos, CA: IEEE Press.
- . 1995. Scheme for Reducing Decoherence in Quantum Computer Memory. *Phys. Rev. A* **52**: 2493.
- . 1997. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM J. Comput.* **26**: 1484.
- Simon, D. R. 1994. On the Power of Quantum Computation. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, p. 116. Los Alamitos, CA: IEEE Press.
- Steane, A. 1996. Multiple Particle Interference and Quantum Error Correction. *Proc. R. Soc. London, Ser. A* **452**: 2551
- Terhal, B. M., and D. P. DiVincenzo. 2000. Problem of Equilibration and the Computation of Correlation Functions on a Quantum Computer. *Phys. Rev. A* **61**: 022301.
- Townsend, P. D. 1998. Quantum Cryptography on Optical Fiber Networks. *Opt. Fiber Tech.: Mat., Dev., Sys.* **4**: 345.
- von Neumann, J. 1932a. Der Messprozess. Ch. VI. In *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer Verlag.
- . 1932b. "Messung und Reversibilität." Allgemeine Betrachtungen. Ch. V. In *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer Verlag.
- Wiesner, S. 1983. Conjugate Coding. *Sigact News* **15**: 78.
- . 1996. Simulations of Many-Body Quantum Systems by a Quantum Computer. [Online]: [http://eprints.lanl.gov. \(quant-ph/9603028\)](http://eprints.lanl.gov. (quant-ph/9603028)).
- Yao, A. 1993. Quantum Circuit Complexity. In *Proceedings of the 34th Annual Symposium on Foundations of Computer Science*. p. 352. Los Alamitos, CA: IEEE Press.
- Zalka, C. 1998. Simulating Quantum Systems on a Quantum Computer. *Proc. R. Soc. London, Ser. A* **454**: 313.

Glossary

- Algorithm.** A set of instructions to be executed by a computing device. What instructions are available depends on the computing device. Typically, instructions include commands for manipulating the contents of memory and means for repeating blocks of instructions indefinitely or until a desired condition is met.
- Amplitude.** A quantum system with a chosen orthonormal basis of "logical" states $|i\rangle$ can be in any superposition $\sum_i \alpha_i |i\rangle$ of these states, where $\sum_i |\alpha_i|^2 = 1$. In such a superposition, the complex numbers α_i are called the amplitudes. Note that the amplitudes depend on the chosen basis.
- Ancillas.** Helper systems used to assist in a computation involving other information systems.
- Bell basis.** For two qubits A and B, the Bell basis consists of the four states $1/\sqrt{2}(|00\rangle_{AB} \pm |11\rangle_{AB})$ and $1/\sqrt{2}(|01\rangle_{AB} \pm |10\rangle_{AB})$.
- Bell states.** The members of the Bell basis.
- Bit.** The basic unit of deterministic information. It is a system that can be in one of two possible states, 0 and 1.

- Bit sequence.** A way of combining bits into a larger system whose constituent bits are in a specific order.
- Bit string.** A sequence of 0s and 1s that represents a state of a bit sequence. Bit strings are the words of a binary alphabet.
- Black box.** A computational operation whose implementation is unknown. Typically, a black box implements one of a restricted set of operations, and the goal is to determine which of these operations it implements by using it with different inputs. Each use of the black box is called a “query.” The smallest number of queries required to determine the operation is called the “query complexity” of the restricted set. Determining the query complexity of sets of operations is an important area of computational complexity.
- Bloch sphere.** The set of pure states of a qubit represented as points on the surface of the unit sphere in three dimensions.
- Bra.** A state expression of the form $\langle\psi|$ considered to be the conjugate transpose of the ket expression $|\psi\rangle$.
- Bra-ket notation.** A way of denoting states and operators of quantum systems with kets (for example, $|\psi\rangle$) and bras (for example, $\langle\phi|$).
- Circuit.** A combination of gates applied to information units in a prescribed order.
To draw circuits, one often uses a convention for connecting and depicting gates. See also “network.”
- Circuit complexity.** The circuit complexity of an operation on a fixed number of information units is the smallest number of gates required to implement the operation.
- Classical information.** The type of information based on bits and bit strings and more generally on words formed from finite alphabets. This is the information used for communication between people. Classical information can refer to deterministic or probabilistic information, depending on the context.
- Computation.** The execution of the instructions provided by an algorithm.
- Computational states.** See “logical states.”
- Computer.** A device that processes information.
- Density matrix or operator.** A representation of pure and mixed states without redundancy. For a pure state $|\psi\rangle$, the corresponding density operator is $|\psi\rangle\langle\psi|$.
A general density operator is a probabilistic combination $\sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$, with $\sum_i \lambda_i = 1$.
- Deterministic information.** The type of information that is based on bits and bit strings.
Deterministic information is classical, but it explicitly excludes probabilistic information.
- Distinguishable states.** In quantum mechanics, two states are considered distinguishable if they are orthogonal. In this case, a measurement exists that is guaranteed to determine which of the two states a system is in.
- Efficient computation.** A computation is efficient if it requires, at most, polynomially many resources as a function of input size. For example, if the computation returns the value $f(x)$ on input x , where x is a bit string, then it is efficient if there exists a power k such that the number of computational steps used to obtain $f(x)$ is bounded by $|x|^k$, where $|x|$ is the length (number of bits) of x .
- Entanglement.** A nonclassical correlation between two quantum systems most strongly exhibited by the maximally entangled states, such as the Bell states for two qubits, and considered to be absent in mixtures of product states (which are called separable states). Often, states that are not separable are considered to be entangled. However, nearly separable states do not exhibit all the features of maximally entangled states. As a result, studies of different types of entanglement are an important component of quantum information theory.
- Gate.** An operation applied to information for the purpose of information processing.

Global phase. Two quantum states are indistinguishable if they differ only by a global phase. That is, $|\psi\rangle$ and $e^{i\phi}|\psi\rangle$ are in essence the same state. The global phase difference is the factor $e^{i\phi}$. The equivalence of the two states is apparent from the fact that their density matrices are the same.

Hilbert space. An n -dimensional Hilbert space consists of all complex n -dimensional vectors. A defining operation in a Hilbert space is the inner product. If the vectors are thought of as column vectors, then the inner product $\langle x, y \rangle$ of x and y is obtained by forming the conjugate transpose x^\dagger of x and calculating $\langle x, y \rangle = x^\dagger y$. The inner product induces the usual squared norm $|x|^2 = \langle x, x \rangle$.

Information. Something that can be recorded, communicated, and computed with. Information is fungible; that is, its meaning can be identified regardless of the particulars of the physical realization. Thus, information in one realization (such as ink on a sheet of paper) can be easily transferred to another (for example, spoken words). Types of information include deterministic, probabilistic, and quantum information. Each type is characterized by information units, which are abstract systems whose states represent the simplest information of each type. The information units define the “natural” representation of the information. For deterministic information, the information unit is the bit, whose states are symbolized by 0 and 1. Information units can be put together to form larger systems and can be processed with basic operations acting on few of them at a time.

Inner product. The defining operation of a Hilbert space. In a finite dimensional Hilbert space with a chosen orthonormal basis $\{e_i : 1 \leq i \leq n\}$, the inner product of two vectors $x = \sum_i x_i e_i$ and $y = \sum_i y_i e_i$ is given by $\sum_i \bar{x}_i y_i$. In the standard column representation of the two vectors, this is the number obtained by computing the product of the conjugate transpose of x with y . For real vectors, that product agrees with the usual “dot” product. The inner product of x and y is often written in the form $\langle x, y \rangle$. Pure quantum states are unit vectors in a Hilbert space. If $|\phi\rangle$ and $|\psi\rangle$ are two quantum states expressed in the ket-bra notation, their inner product is given by $(|\phi\rangle)^\dagger \langle \psi| = \langle \phi|\psi\rangle$.

Ket. A state expression of the form $|\psi\rangle$ representing a quantum state. Usually, $|\psi\rangle$ is thought of as a superposition of members of a logical state basis $|i\rangle$. One way to think about the notation is to consider the two symbols $|$ and \rangle as delimiters denoting a quantum system and ψ as a symbol representing a state in a standard Hilbert space. The combination $|\psi\rangle$ is the state of the quantum system associated with ψ in the standard Hilbert space via a fixed isomorphism. In other words, one can think of $\psi \leftrightarrow |\psi\rangle$ as an identification of the quantum system’s state space with the standard Hilbert space.

Linear extension of an operator. The unique linear operator that implements a map defined on a basis. Typically, we define an operator U on a quantum system only on the logical states $U : |i\rangle \rightarrow |\psi_i\rangle$. The linear extension is defined by $U(\sum_i \alpha_i |i\rangle) = \sum_i \alpha_i |\psi_i\rangle$.

Logical states. For quantum systems used in information processing, the logical states are a fixed orthonormal basis of pure states. By convention, the logical basis for qubits consists of $|0\rangle$ and $|1\rangle$. For larger dimensional quantum systems, the logical basis is often indexed by integers, $|0\rangle, |1\rangle, |2\rangle$, and so on. The logical basis is often called the computational basis, or sometimes, the classical basis.

Measurement. The process used to extract classical information from a quantum system. A general projective measurement is defined by a set of projectors P_i , satisfying $\sum_i P_i = \mathbb{1}$ and $P_i P_j = \delta_{ij} P_i$. Given the quantum state $|\psi\rangle$, the outcome of a measurement with the set $\{P_i\}_i$ is one of the classical indices i associated with a projector P_i . The index i is the measurement outcome. The probability of outcome i is $p_i = |P_i |\psi\rangle|^2$, and given outcome i , the quantum state “collapses” to $P_i |\psi\rangle / \sqrt{p_i}$.

Mixture. A probabilistic combination of the pure states of a quantum system. Mixtures can be represented without redundancy with density operators. Thus, a mixture is of the form $\sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$, with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$ being the probabilities of the states $|\psi_i\rangle$. This expression for mixtures defines the set of density operators, which can also be characterized as the set of operators ρ satisfying $\text{tr}(\rho) = 1$ and for all $|\psi\rangle$, $\langle\psi|\rho|\psi\rangle \geq 0$ (“positive semidefinite operator”).

Network. In the context of information processing, a network is a sequence of gates applied to specified information units. Networks can be visualized as displaying horizontal lines that denote the timeline of an information unit. The gates are represented by graphical elements that intercept the lines at specific points. A realization of the network requires applying the gates to the information units in the specified order (left to right).

Operator. A function that transforms the states of a system. Operators may be restricted depending on the system’s properties. For example, in talking about operators acting on quantum systems, one always assumes that they are linear.

Oracle. An information processing operation that can be applied. A use of the oracle is called a query. In the oracle model of computation, a standard model is extended to include the ability to query an oracle. Each oracle query is assumed to take one time unit. Queries can reduce the resources required for solving problems. Usually, the oracle implements a function or solves a problem not efficiently implementable by the model without the oracle. Oracle models are used to compare the power of two models of computation when the oracle can be defined for both models. In 1994, for example, Dan Simon showed that quantum computers with a specific oracle O could efficiently solve a problem that had no efficient solution on classical computers with access to the classical version of O . At the time, this result was considered the strongest evidence for an exponential gap in power between classical and quantum computers.

Overlap. The inner product between two quantum states.

Pauli operators. The Hermitian matrices σ_x , σ_y , and σ_z acting on qubits, which are two-level quantum systems. They are defined in Equation (12). It is often convenient to consider the identity operator to be included in the set of Pauli operators.

Polynomial resources. To say that an algorithm computing the function $f(x)$, where x is a bit string, uses polynomial resources (in other words, is efficient) means that the number of steps required to compute $f(x)$ is bounded by $|x|^k$ for some fixed k . Here, $|x|$ denotes the length of the bit string x .

Probabilistic bit. The basic unit of probabilistic information whose state space consists of all probability distributions over the two states of a bit. The states can be thought of as describing the outcome of a biased coin flip before the coin is flipped.

Probabilistic information. The type of information obtained by extending the state spaces of deterministic information to allow arbitrary probability distributions over the deterministic states. This is the main type of classical information with which quantum information is compared.

Probability amplitude. The squared norm of an amplitude with respect to a chosen orthonormal basis $\{|i\rangle\}$. Thus, the probability amplitude is the probability with which the state $|i\rangle$ is measured in a complete measurement that uses this basis.

Product state. For two quantum systems A and B, product states are of the form $|\psi\rangle_A |\phi\rangle_B$. Most states are not of this form.

Program. An algorithm expressed in a language that can be understood by a particular type of computer.

Projection operator. A linear operator P on a Hilbert space that satisfies $P^2 = P^\dagger P = P$. The projection onto a subspace V with orthogonal complement W is defined as follows: If $x \in V$ and $y \in W$, then $P(x + y) = x$.

- Pseudocode.** A semiformal computer language intended to be executed by a standard random-access machine, which is a machine model with a central processing unit and access to a numerically indexed unbounded memory. This machine model is representative of the typical one-processor computer. Pseudocode is similar to programming languages such as BASIC, Pascal, or C but does not have specialized instructions for human interfaces, file management, or other “external” devices. Its main use is to describe algorithms and enable machine-independent analysis of the algorithms’ resource usage.
- Pure state.** A state of a quantum system that corresponds to a unit vector in the Hilbert space used to represent the system’s state space. In the ket notation, pure states are written in the form $|\psi\rangle = \sum_i \alpha_i |i\rangle$, where the $|i\rangle$ form a logical basis and $\sum_i |\alpha_i|^2 = 1$.
- Quantum information.** The type of information obtained when the state space of deterministic information is extended by normalized superpositions of deterministic states. Formally, each deterministic state is identified with one of an orthonormal basis vector in a Hilbert space, and normalized superpositions are unit-length vectors expressible as complex linear sums of the chosen basis vectors. It is convenient to extend this state space further by permitting probability distributions over the quantum states (see the entry for “mixtures”). This extension is still called quantum information.
- Qubit.** The basic unit of quantum information. It is the quantum extension of the deterministic bit, which implies that its state space consists of the unit-length vectors in a two-dimensional Hilbert space.
- Readout.** A method for obtaining human-readable information from the state of a computer. For quantum computers, readout refers to a measurement process used to obtain classical information about a quantum system.
- Reversible gate.** A gate whose action can be undone by a sequence of gates.
- Separable state.** A mixture of product states.
- States.** The set of states for a system characterizes the system’s behavior and possible configurations.
- Subspace.** For a Hilbert space, a subspace is a linearly closed subset of the vector space. The term can be used more generally for a system Q of any information type: A subspace of Q or, more specifically, of the state space of Q is a subset of the state space that preserves the properties of the information type represented by Q .
- Superposition principle.** One of the defining postulates of quantum mechanics according to which if states $|1\rangle, |2\rangle, \dots$ are distinguishable, then $\sum_i \alpha_i |i\rangle$ with $\sum_i |\alpha_i|^2 = 1$ is a valid quantum state. Such a linear combination is called a normalized superposition of the states $|i\rangle$.
- System.** An entity that can be in any of a specified number of states. An example is a desktop computer whose states are determined by the contents of its various memories and disks. Another example is a qubit, which can be thought of as a particle whose state space is identified with complex, two-dimensional, length-one vectors. Here, a system is always associated with a type of information that determines the properties of the state space. For example, for quantum information, the state space is a Hilbert space. For deterministic information, it is a finite set called an alphabet.
- Unitary operator.** A linear operator U on a Hilbert space that preserves the inner product. That is, $\langle Ux, Uy \rangle = \langle x, y \rangle$. If U is given in matrix form, then this expression is equivalent to $U^\dagger U = \mathbb{1}$.
- Universal set of gates.** A set of gates that satisfies the requirement that every allowed operation on information units can be implemented by a network of these gates. For quantum information, it means a set of gates that can be used to implement every unitary operator. More generally, a set of gates is considered universal if, for every operator U , there are implementable operators V arbitrarily close to U .

From Factoring to Phase Estimation

A discussion of Shor's algorithm

Emanuel Knill, Raymond Laflamme, Howard N. Barnum, Diego A. Dalvit, Jacek J. Dziarmaga, James E. Gubernatis, Leonid Gurvits, Gerardo Ortiz, Lorenza Viola, and Wojciech H. Zurek

The publication of Shor's quantum algorithm for efficiently factoring numbers (1994 and 1997) was the key event that stimulated many theoretical and experimental investigations of quantum computation. One of the reasons why this algorithm is so important is that the security of widely used public-key cryptographic protocols relies on the conjectured difficulty of factoring large numbers. An elementary overview of these protocols and the quantum algorithm for breaking them is provided in Artur Ekert (1998).¹ Here, we outline the relationship between factoring and the powerful technique of phase estimation. This relationship helps in understanding many of the existing quantum algorithms and was first explained in Richard Cleve et al. (1998). This explanation was motivated by Alexei Kitaev's version (1995) of the factoring algorithm.

The factoring problem requires writing a whole number N as a product of primes. (Primes are whole numbers greater than 1 that are divisible without remainder only by 1 and themselves.) Shor's algorithm solves this problem by reducing it to instances of the order-finding problem, which will be defined below. The reduction is based on basic number theory and involves efficient classical computation. At the core of Shor's algorithm is a quantum algorithm that solves the order-finding problem efficiently. In this case, an algorithm is considered efficient if it uses resources bounded by a polynomial in the number of digits of N . For more information on the requisite number theory, see any textbook on number theory (Bolker 1970, Hardy and Wright 1979).

We begin by showing that factoring reduces to order finding. The first observation is that, to factor a whole number, it is sufficient to solve the factor-finding problem, whose statement is, "Given a whole number N , find a proper factor of N if one exists. A factor of N is a whole number f that satisfies $N = fg$ for some whole number g . The factor f is proper if $f \neq 1$ and $f \neq N$. For example, if $N = 15$, then 3 and 5 are its proper factors. For some numbers, it is easy to find proper factors. For example, you can tell that N is even from the least significant digit (in decimal or binary), in which case, 2 is a proper factor (unless $N = 2$, a prime). But many numbers are not so easy. As an example, you can try to find a factor of $N = 149,573$ by hand.² You can complete the factorization of a whole number by recursively applying an algorithm for the factor-finding problem to all the proper factors found.

Before we continue the reduction of factoring to order finding, we will briefly explain modular arithmetic, which both simplifies the discussion and is necessary to avoid computing with numbers that have exponential numbers of digits. We say that a and b are equal modulo N , written as $a = b \pmod{N}$, if $a - b$ is divisible by N (without remainder). For example, $3 = 18 \pmod{15} = 33 \pmod{15}$. Equality modulo N is well behaved with respect to addition and multiplication. That is, if $a = b \pmod{N}$ and $c = d \pmod{N}$, then $a + c = b + d \pmod{N}$, and $ac = bd \pmod{N}$. For factoring N , we will be look-

¹All the citations in this article have been referenced on pages 31 to 33 of the main article, "Quantum Information Processing."

²149573 * 373 = 54957361

ing for whole numbers a that are divisible by a proper factor of N . If a has this property, then so does any b with $b = a \pmod N$. We therefore perform all arithmetic modulo N .

One way to think of all this is that we use only whole numbers a that satisfy $0 \leq a \leq N - 1$. We can implement each arithmetic operation modulo N by applying the operation in the usual way and then computing the remainder after division by N . For example, to obtain $ab \pmod N$, we first compute ab . The unique c such that $0 \leq c \leq N - 1$ and $c = ab \pmod N$ is the remainder after division of ab by N . Thus, c is the result of multiplying a by b modulo N . Consistent with this procedure, we can think of the expression $a \pmod N$ as referring to the remainder of a after division by N .

The second observation in the reduction of factoring to order finding is that it is sufficient to find a whole number r with the property that $r^2 - 1$ is a multiple of N , but $r - 1$ and $r + 1$ are not. Using the language of modular arithmetic, the property is expressed as $r^2 = 1 \pmod N$, but $r \neq 1 \pmod N$ and $r \neq -1 \pmod N$. Because $1 \pmod N$ and $-1 \pmod N$ are the obvious square roots of $1 \pmod N$, we say that r is a nontrivial square root of unity (modulo N). For such an r , one can write $r^2 - 1 = (r - 1)(r + 1) = mN$ for some whole number m . This implies that every prime factor p of N divides either $(r - 1)$ or $(r + 1)$ so that either $(r - 1)$ or $(r + 1)$ is or shares a factor with N . Suppose that $r - 1$ is or shares such a factor. Because $r - 1$ is not a multiple of N , the greatest common divisor of $r - 1$ and N is a factor of N . Since an efficient classical algorithm (the Euclidean algorithm) exists for finding the greatest common divisor, we can easily find the desired proper factor.

The examples of $N = 15$ and $N = 21$ serve to illustrate the key features of the algorithm. For $N = 15$, possible choices for r are $r = 4$ ($4^2 - 1 = 1 * 15$), and $r = 11$ ($11^2 - 1 = 120 = 8 * 15$). For the first choice, the proper factors emerge immediately: $4 - 1 = 3$, and $4 + 1 = 5$. For the second, it is necessary to determine the greatest common divisors (or gcd). Let $\text{gcd}(x, y)$ stand for the greatest common divisor of x and y . The proper factors are $\text{gcd}(11 - 1, 15) = \text{gcd}(10, 15) = 5$, and $\text{gcd}(11 + 1, 15) = \text{gcd}(12, 15) = 3$. For $N = 21$, one can take $r = 8$ as $8^2 - 1 = 63 = 3 * 21$. In this case, $8 - 1 = 7$ is a proper factor, and $\text{gcd}(8 + 1, 21) = 3$ is another.

For N even or a power of a prime, it is not always possible to find a nontrivial square root of unity. Because both cases can be handled efficiently by known classical algorithms, we can exclude them. In every other case, such numbers r exist. One way to find such an r is to start from any whole number q , with $1 < q < N$. If $\text{gcd}(q, N) = 1$, then according to a basic result in number theory, there is a smallest whole number $k > 1$ such that $q^k - 1 = 0 \pmod N$. The number k is called the order of q modulo N . If k is even, say, $k = 2l$, then $(q^l)^2 = 1 \pmod N$, so q^l is a (possibly trivial) square root of unity. For the example of $N = 15$, we can try $q = 2$. The order of 2 modulo 15 is 4, which gives $r = 2^2 = 4$, the first of the two choices in the previous paragraph. For $N = 21$, again with $q = 2$, the order is 6: $2^6 - 1 = 63 = 3 * 21$. Thus, $r = 2^3 = 8$. We can also try $q = 11$, in which case, with foresight, it turns out that $11^6 - 1$ is divisible by 21. A possible problem appears, namely, the powers q^k , which we want to compute, are extremely large. But modular arithmetic can help us avoid this problem. For example, to find the order of 11 modulo 21 by direct search, we can perform the following computation: In general, such a direct search for the order of q modulo N is very inefficient, but as we will see,

$$\begin{aligned}
 11^2 &= 121 = 5 * 21 + 16 = 16 \pmod{21} \\
 11^3 &= 11 * 11^2 = 11 * 16 \pmod{21} = 11 * (-5) \pmod{21} \\
 &= -55 \pmod{21} = -3 * 21 + 8 \pmod{21} = 8 \pmod{21} \\
 11^4 &= 11 * 11^3 = 11 * 8 \pmod{21} = 4 * 21 + 4 \pmod{21} = 4 \pmod{21} \\
 11^5 &= 11 * 11^4 = 11 * 4 \pmod{21} = 2 \pmod{21} \\
 11^6 &= 11 * 11^5 = 11 * 2 \pmod{21} = 1 \pmod{21}
 \end{aligned} \tag{1}$$

there is an efficient quantum algorithm that can determine the order.

A factor-finding algorithm based on the above observations is the following:

FACTORFIND(N)

Input: A positive, nonprime whole number N

Output: A proper factor f of N , that is, f is a whole number such that $1 < f < N$ and $N = fg$ for some whole number g .

1. If N is even, return $f = 2$.
2. If $N = p^k$ for p prime, return p .
3. Randomly pick $1 < q < N - 1$.
 - a. If $f = \gcd(q, N) > 1$, return f .
4. Determine the order k of q modulo N using the quantum order-finding algorithm.
 - a. If k is not even, repeat at step 3.
5. Write $k = 2l$ and determine $r = q^l \bmod N$ with $1 < r < N$.
 - a. If $1 < f = \gcd(r - 1, N) < N$, return f .
 - b. If $1 < f = \gcd(r + 1, N) < N$, return f .
 - c. If we failed to find a proper factor, repeat at step 3.

The efficiency of this algorithm depends on the probability that a randomly chosen q at step 3 results in finding a factor. An analysis of the group of numbers q that satisfy $\gcd(q, N) = 1$ shows that this probability is sufficiently large.

The main problem left to be solved is finding the order of $q \bmod N$. A direct search for the order of $q \bmod N$ involves computing the sequence

$$1 \rightarrow q \rightarrow q^2 \bmod N \rightarrow \dots \rightarrow q^{k-1} \bmod N \rightarrow 1 = q^k \bmod N . \quad (2)$$

This sequence can be conveniently visualized as a cycle whose length is the order $q \bmod N$ (refer to Figure 1).

To introduce the quantum algorithm, we first associate the logical quantum states $|0\rangle, |1\rangle, \dots, |N-1\rangle$ with the numbers $0, 1, \dots, N-1$. The map f that takes each number on the cycle to the next number along the cycle is given by $f(x) = qx \bmod N$. For q satisfying $\gcd(q, N) = 1$, the map f permutes not only the numbers on the cycle but all the numbers modulo N . As a result, the linear operator \hat{f} defined by $\hat{f}|x\rangle = |f(x)\rangle = |qx \bmod N\rangle$ is unitary. The quantum algorithm deduces the length of the cycle for q by making measurements to determine the properties of the action of \hat{f} on superpositions of the states $|q^s \bmod N\rangle$. To illustrate the basic ideas, we work out the example of $N = 15$ and $q = 8$. The action of \hat{f} on the states $|1\rangle, |8\rangle, |4\rangle$, and $|2\rangle$ in the cycle of $8 \bmod 15$ is

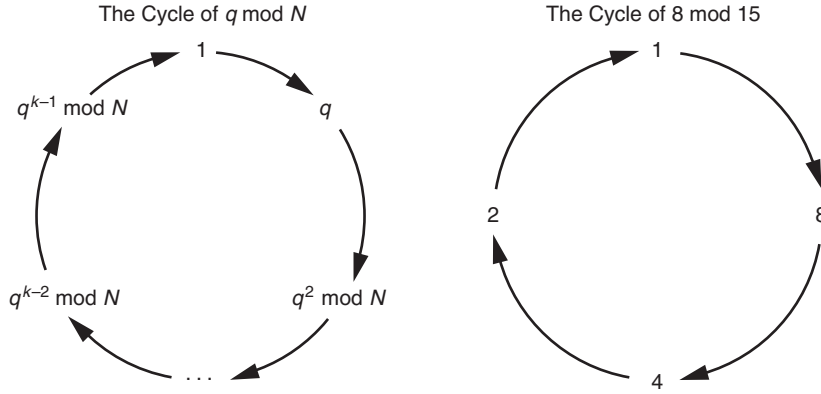


Figure 1. Multiplicative Cycles of $q \bmod N$
 Each number on a cycle is obtained from the previous one by multiplication by $q \bmod N$.

completely determined by the eigenstates and eigenvalues of \hat{f} . For cyclicly acting permutations, a basis of eigenstates is given by the Fourier basis for the space spanned by the states in a cycle. For the cycle of interest, the Fourier basis consists of the states

$$\begin{aligned}
 |\psi_0\rangle &= \frac{1}{2}(|1\rangle + |8\rangle + |4\rangle + |2\rangle) , \\
 |\psi_1\rangle &= \frac{1}{2}(|1\rangle + i|8\rangle - |4\rangle - i|2\rangle) , \\
 |\psi_2\rangle &= \frac{1}{2}(|1\rangle - |8\rangle + |4\rangle - |2\rangle) , \text{ and} \\
 |\psi_3\rangle &= \frac{1}{2}(|1\rangle - i|8\rangle - |4\rangle + i|2\rangle) .
 \end{aligned} \tag{3}$$

The phases of the l^{th} state of the cycle occurring in the sum for $|\psi_m\rangle$ can be written as i^{lm} . It follows that $\hat{f}|\psi_m\rangle = i^m|\psi_m\rangle$, that is, the eigenvalue of \hat{f} for $|\psi_m\rangle$ is i^m . Note that, in complex numbers, the powers of i are all the fourth roots of unity. In general, the Fourier basis for the cycle $\dots \rightarrow |q^l \bmod N\rangle \rightarrow \dots$ consists of the states $|\psi_m\rangle = \sum_l \omega^{lm}|q^l \bmod N\rangle$, where $\omega = e^{i2\pi/k}$ is a primitive k^{th} root of unity. (The complex number x is a primitive k^{th} root of unity if k is the smallest whole number $k > 0$ such that $x^k = 1$. For example, both -1 and i are fourth roots of unity, but only i is primitive.)

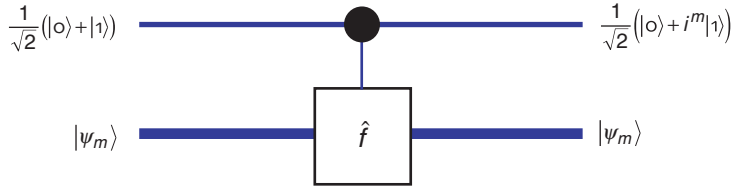
It is, of course, possible to express the logical state $|1\rangle$ using the Fourier basis

$$|1\rangle = \frac{1}{2}(|\psi_0\rangle + |\psi_1\rangle + |\psi_2\rangle + |\psi_3\rangle) . \tag{4}$$

The key step of the quantum algorithm for order finding consists of a measurement to estimate a random eigenvalue of \hat{f} , whose associated eigenstate occurs in the expression for $|1\rangle$ in terms of the Fourier basis. If the eigenvalue found is a k^{th} root of unity, we infer that the cycle length is divisible by k and check (using a classical algorithm) whether this is the order of q . In the example, the random eigenvalues are 1 (the only primitive first root of unity), i and $-i$ (primitive fourth roots of unity), and -1 (the primitive second root of unity). The order is found if the random eigenvalue is a fourth root of unity, which happens with probability 1/2 in this case.

Figure 2. Phase Estimation with One Qubit

The input is a product state on one ancilla qubit and on a second quantum system, as shown. The state $|\psi_m\rangle$ on the second system is an eigenstate of \hat{f} . For the example provided in Equation (3), the eigenvalue is i^m . A controlled- \hat{f} operation is applied to the input, that is, \hat{f} is applied to the second system conditional on $|1\rangle$ for the ancilla qubit. In the bra-ket notation, the total operation can be written as $|0\rangle\langle 0| + |1\rangle\langle 1|\hat{f}$ (system labels have been omitted). Because \hat{f} changes only the phase of its input, the second system is unchanged, but the phase modifies the ancilla qubit's superposition as shown.



The quantum algorithm for obtaining an eigenvalue is called the phase estimation algorithm, and it exploits a more general version of the phase kickback we encountered in the solution of the parity problem. The phase kickback transfers the eigenvalue of an eigenstate of \hat{f} to a Fourier basis on a number of additional qubits called helper or ancilla qubits. Which Fourier state results is then determined by a subroutine called the measured quantum Fourier transform. We introduce these elements in the next paragraphs. Their combination for solving the general order-finding problem is illustrated on page 45.

Figure 2 shows how to kick back the eigenvalue of an eigenstate of \hat{f} using a network implementing the controlled- \hat{f} operation. The network in Figure 2 can be used with input $|1\rangle$ on the second system. From Equation (4) and the superposition principle, it follows that the output correlates the different phase kickback states with the four eigenvectors $|\psi_m\rangle$. That is, the network implements the following transformation:

$$\frac{1}{2\sqrt{2}} (|0\rangle + |1\rangle) \begin{pmatrix} |\psi_0\rangle \\ + |\psi_1\rangle \\ + |\psi_2\rangle \\ + |\psi_3\rangle \end{pmatrix} \rightarrow \frac{1}{2\sqrt{2}} \begin{pmatrix} (|0\rangle + i^0|1\rangle) |\psi_0\rangle \\ + (|0\rangle + i^1|1\rangle) |\psi_1\rangle \\ + (|0\rangle + i^2|1\rangle) |\psi_2\rangle \\ + (|0\rangle + i^3|1\rangle) |\psi_3\rangle \end{pmatrix}. \quad (5)$$

The hope is that a measurement of the first qubit can distinguish between the four possible phases that can be kicked back. However, because the four states are not mutually orthogonal, they are not unambiguously distinguishable by a measurement. To solve this problem, we use a second qubit and a controlled- \hat{f}^2 as shown in Figure 3.

The four possible states $|u_m\rangle$ that appear on the ancilla qubits in the network of Figure 3 are the Fourier basis for the cycle $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 0$ and are therefore orthonormal. If we apply the network of Figure 3 with $|1\rangle$ instead of $|\psi_m\rangle$ at the lower input, the output correlates the four $|\psi_m\rangle$ in the superposition with the $|u_m\rangle$, which makes the information about the eigenvalues of \hat{f} available in the Fourier basis of the two ancilla qubits. This approach has the advantage that the states are known, whereas in the Fourier basis for the cycle of $q \bmod N$, the states depend on the numbers in the cycle, which are not known in advance (except in very simple cases, such as the example we are working with).

To learn one of the eigenvalues of \hat{f} , the last step is to make a measurement in the Fourier basis. For one qubit representing the binary numbers 0 and 1, the Fourier basis is $1/\sqrt{2}(|0\rangle + |1\rangle)$ and $1/\sqrt{2}(|0\rangle - |1\rangle)$, which is constructed as discussed after

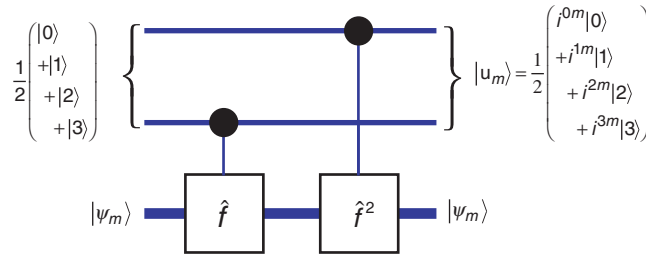


Figure 3. Phase Estimation with Two Qubits

Using two qubits ensures distinguishability of the eigenvalues of \hat{f} for the states $|\psi_m\rangle$. The states of the input qubits are used to represent the numbers from 0 to 3 in binary. The most significant bit (the two's digit in binary representation) is carried by the top qubit. That is, we make the following identification: $|0\rangle = |00\rangle$, $|1\rangle = |01\rangle$, $|2\rangle = |10\rangle$, and $|3\rangle = |11\rangle$. It follows that the network has the effect of applying \hat{f}^m conditional on the input qubits' logical state being $|m\rangle$.

Equation (3) but using the square root of unity $\omega = -1$ instead of the fourth root i . To make a measurement that determines which of the two basis vectors is present, it suffices to apply the Hadamard transform \mathbf{H} and make a standard measurement, just as we did twice in the network of Figure 2 in the article “Quantum Information Processing” on page 23. A more complicated network works with two qubits representing the binary numbers from 0 to 3. Such a network is shown in Figure 4.

To see how the network extracts the bits in the index of $|u_a\rangle$, we can follow the states as the network is executed. The input state at checkpoint 1 in Figure 4 is given by

$$|\phi_1\rangle = |u_a\rangle = \frac{1}{2} \begin{pmatrix} i^{0*a}|0\rangle \\ +i^{1*a}|1\rangle \\ +i^{2*a}|2\rangle \\ +i^{3*a}|3\rangle \end{pmatrix} = \frac{1}{2} \begin{pmatrix} i^{(0*2^1+0*2^0)}(a_1*2^1+a_0*2^0)|00\rangle \\ +i^{(0*2^1+1*2^0)}(a_1*2^1+a_0*2^0)|01\rangle \\ +i^{(1*2^1+0*2^0)}(a_1*2^1+a_0*2^0)|10\rangle \\ +i^{(1*2^1+1*2^0)}(a_1*2^1+a_0*2^0)|11\rangle \end{pmatrix}. \tag{6}$$

In the last sum, the relevant numbers have been fully expanded in terms of their binary digits to give a flavor of the general principles underlying the measured Fourier transform. The next step of the network applies a Hadamard gate to the qubit carrying the most significant digit. To understand how it succeeds in extracting a_0 , the least signifi-

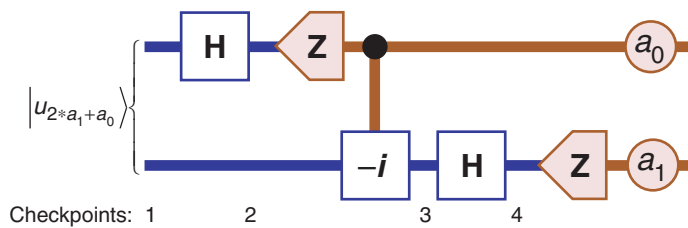


Figure 4. Measured Quantum Fourier Transform on Two Qubits
The two qubits represent the numbers 0, 1, 2, and 3. If the input is one of the Fourier states $|u_a\rangle$, where the binary digits of a are determined by $a = 2 * a_1 + a_0$, then the measurement outcomes are a_0 and a_1 , as shown. The numbers under the network are checkpoints used for analysis. [For details on the measured Fourier transform, see Griffiths and Niu (1996).]

cant bit of a , let b with binary digits b_0 and b_1 represent one of the logical states of the two qubits. As before, the most significant bit b_1 is represented by the top/first qubit that the first Hadamard gate is applied to. The phase of $|b\rangle$ in Equation (6) is given by $i^{(b_1*2^1+b_0*2^0)(a_1*2^1+a_0*2^0)}$. Next, we determine how the phase depends on b_1 :

$$\begin{aligned}
 i^{(b_1*2^1+b_0*2^0)(a_1*2^1+a_0*2^0)} &= i^{b_1*2^1*(a_1*2^1+a_0*2^0)} i^{b_0*2^0*(a_1*2^1+a_0*2^0)} \\
 &= i^{b_1*a_1*2^2} i^{b_1*a_0*2^1} i^{b_0*2^0*(a_1*2^1+a_0*2^0)} \\
 &= (i^4)^{b_1*a_1} (i^2)^{b_1*a_0} i^{b_0*2^0*(a_1*2^1+a_0*2^0)} \\
 &= (-1)^{b_1*a_0} i^{b_0*2^0*(a_1*2^1+a_0*2^0)} .
 \end{aligned} \tag{7}$$

It follows that, if $a_0 = 0$, the phase does not depend on b_1 , and if $a_0 = 1$, it changes sign with b_1 . This sign change can be detected by performing the Hadamard transform and measuring, as can be seen explicitly by computing the state after the Hadamard transform at checkpoint 2:

$$\begin{aligned}
 |\phi_2\rangle &= \frac{1}{\sqrt{2}} \left(i^{0*2^0*(a_1*2^1+a_0*2^0)} |a_0\rangle |0\rangle + i^{1*2^0*(a_1*2^1+a_0*2^0)} |a_0\rangle |1\rangle \right) \\
 &= |a_0\rangle \frac{1}{\sqrt{2}} \left(i^{0*2^0*(a_1*2^1+a_0*2^0)} |0\rangle + i^{1*2^0*(a_1*2^1+a_0*2^0)} |1\rangle \right) .
 \end{aligned} \tag{8}$$

The phases still show a dependence on a_0 via the terms $i^{b_0*2^0*a_0*2^0} = i^{b_0a_0}$. The purpose of the phase-shift gate conditioned on the measurement outcome is to remove that dependence. The result is the following state on the remaining qubit at checkpoint 3:

$$\begin{aligned}
 |\phi_3\rangle &= \frac{1}{\sqrt{2}} \left(i^{0*2^0*a_1*2^1} |0\rangle + i^{1*2^0*a_1*2^1} |1\rangle \right) \\
 &= \frac{1}{\sqrt{2}} \left((-1)^{0*a_1} |0\rangle + (-1)^{1*a_1} |1\rangle \right) \\
 &= \frac{1}{\sqrt{2}} \left(|0\rangle + (-1)^{a_1} |1\rangle \right) .
 \end{aligned} \tag{9}$$

The final Hadamard transform followed by a measurement therefore results in the bit a_1 , as desired.

The elements that we used to determine the order of 8 modulo 15 can be combined and generalized to determine the order of any q modulo N with $\gcd(q, N) = 1$. The general network is shown in Figure 5. Two features of the generalization are not apparent from the example. First, in order for the quantum network to be efficient, an efficient implementation of the controlled \hat{f}^{2^l} operation is required. To obtain such an implementation, first note that to calculate $\hat{f}^{2^l}(x) = q^{2^l} x \bmod N$, it suffices to square q repeatedly

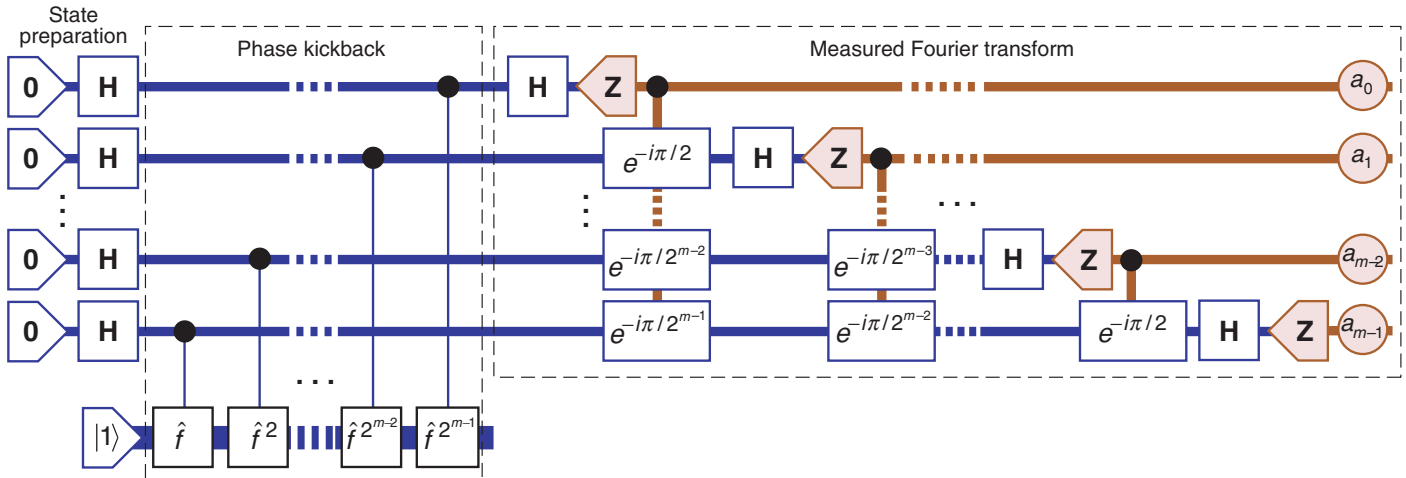


Figure 5. Network for Quantum Order Finding and Phase Estimation

The number m of qubits used for the phase kickback has to be chosen such that $m > 2 * \log_2(k_u)$, where k_u is a known upper bound on the order k of $q \bmod N$. Because $N > k$, one can set $m = 2 \lceil \log_2(N) \rceil$, where $\lceil x \rceil$ is the least whole number $s \geq x$. There is an eigenvalue $\lambda_l = e^{i2l\pi/k}$ of one of the Fourier eigenvectors associated with the cycle of $q \bmod N$ such that the number a , whose binary digits are the measurement outcomes, satisfies $e^{i\pi a/2^{m-1}} \approx e^{i2\pi l/k}$. More precisely, with probability above .405, there exists l such that $|a/2^m - l/k| \leq 1/2^{m+1}$ (Cleve et al. 1998). Because any two distinct rational numbers with denominator at most k_u differ by at least $1/k_u^2 > 1/2^{m+1}$, the theory of rational

approximations guarantees that we can uniquely determine the number l/k . There is an efficient classical algorithm based on continued fractions that computes r and s with $r/s = l/k$ and $s = k/\text{gcd}(l, k)$. The probability that $\text{gcd}(l, k) = 1$ is at least $1/(\log_2(k) + 1)$, in which case we learn that $s = k$ and this is the order of $q \bmod N$. Note that the complexity of the network depends on the complexity of implementing the controlled \hat{f}^{2^l} operations. Because these operations can be implemented efficiently, the network and hence the determination of the order of $q \bmod N$ are efficient in the sense that, on average, polynomial resources in $\log_2(N)$ suffice.

modulo N using $(q^{2^m})^2 \bmod N = q^{2^{m+1}} \bmod N$ until we obtain $q^{2^l} \bmod N$. The result is then multiplied by $x \bmod N$. This computation is efficient. For any given q , the computation can be converted to an efficient network consisting of Toffoli gates and controlled-not gates acting on the binary representation of x . The conversion can be accomplished with standard techniques from the theory of reversible classical computation. The result is an efficient network for \hat{f}^{2^l} . Basic network theory can then be used to implement the controlled version of this operation (Barenco et al. 1995).

To understand the second feature, note that we were lucky to anticipate that the order of 8 modulo 15 was a power of 2, which nicely matched the measured Fourier transform we constructed on two qubits. The measured Fourier transform on m ancilla qubits can detect exactly only eigenvalues that are powers of the 2^m th root of unity $e^{i\pi/2^{m-1}}$. The phase kicked back by the controlled operations corresponds to a k^{th} root of unity. Given a Fourier state on the cycle of $q \bmod N$, the resulting state on the ancilla qubits has phases that go as powers of a k^{th} root of unity. Fortunately, the ancilla's Fourier basis is such that the measured Fourier transform picks up primarily those basis states whose generating phase is close to the kickback phase. Thus, we are likely to detect a nearby $\omega = e^{i\pi/2^{m-1}}$. It is still necessary to infer (a divisor of) k from knowledge of such an ω . Because we know that the order k is bounded by N , the number of possible phases kicked back that are near the measured ω is limited. To ensure that there is only one possible such phase, it is necessary to choose m such that $2^m > N^2$. (See also Figure 5.) ■



cryptology
quantum computers
cryptology
quantum computers

QUESTIONS, Quantum Computers, and Cryptography

A mathematical metaphor for the power of quantum algorithms

Mark Ettinger

How can quantum computers do the amazing things that they are able to do, such as factoring large numbers and finding discrete logarithms? What makes them so different from classical computers? These questions are often asked, and they have proved to be surprisingly difficult to answer—at least to the satisfaction of everyone! In this short article, I'll try to address these questions by comparing the operation of a quantum computer with playing the game of 20 questions. But first, let's consider an unusual perspective on computers in general.

What Is a Computer?

Well, a computer is really just some physical machine that you prepare in a certain way, manipulate in certain ways, and then watch to observe the results it displays. That is how physicists might describe the entire physical process that mathematicians call a computation. This view seems a bit strange at first because we have become accustomed to the more abstract view of the computer scientist, who sees a computation as a certain type of process that acts on an input in order to produce an output. But our physical description is not really so different. It just emphasizes the physical nature of the computation, something that falls by the wayside in the abstracted view. The initial preparation is what a computer scientist calls an input, the actual computation is the physical manipulation, and the observation at the end results in getting the output. So, whereas a computation can be viewed abstractly as a process, its physical nature can also be emphasized. This view will help us make the transition to understanding what a quantum machine is doing in a special way. Unlike classical computers, which are physical devices manipulated according to the laws of classical physics, quantum computers are physical devices manipulated according to the laws of quantum physics.

Quantum Computers and the 20 Questions Game

Having understood that a computation is ultimately a physical process, let's go on to see how using a quantum machine is much like playing the game of 20 questions. Twenty questions is played as follows. I think of a number between 1 and 2^{20} . You try to guess my secret number by asking questions such as, "Is your secret number less than 2378?" If you ask your questions well, you can guess my secret number in, at the most, 20 questions. Why? Well, with each question, you can eliminate half of the remaining candidates. Computer scientists call this process binary search, and it allows you to find a secret number less than 2^n in $\log 2^n = n$ questions at the most. The key idea is that, by cutting the number of possibilities in half with each question, you are left with one possibility after only n questions. This principle generalizes. For example, if you are searching for a secret item among N possibilities and with each question you are able to eliminate a fraction $1 - 1/c$ of the possibilities, then you can find the secret in $\log_c N$ questions. In general, you might not be looking for a number. You might be looking for a secret element x in a set S called a search space. The key to quick success is still to be able to eliminate a constant fraction of the remaining candidates. Now, let's consider a slightly different version of this game, which we call "random 20 questions."

In playing random 20 questions, you don't get to choose your question. Instead, you randomly select a subset Q (used for the word "question") consisting of half of the N elements in the search space, and you ask, "Is the secret element in Q ?" After I give you the honest answer, you choose a new random subset Q and ask again. Surprisingly, again after only about $\log N$ questions, you will almost surely have narrowed the possibilities down to the one correct answer. We say "almost surely" because there is a tiny, tiny chance that you will get unlucky and never be able to eliminate one of the elements that is not the secret element. This tiny chance is the result of each question having been selected randomly rather than deterministically, which is the case when playing the original 20 questions game. After $2 \log N$ questions, for example, that possibility is incredibly small. So, even by asking random questions, you can discover the secret element quickly. The reason is that, as in the original 20 questions game, you are able to eliminate each incorrect element as a possibility. Although in the random 20 questions game this process of elimination is only very highly probable, it is so close to being certain that, for all practical purposes, we won't worry about it. Now, let's talk about playing quantum 20 questions.

In this game, I choose a secret quantum state ρ_1 from a search space of quantum states $S = \{\rho_1, \rho_2, \dots, \rho_N\}$, and I supply a copy of the secret state whenever you request one. Your task is to discover my secret quantum state by asking quantum questions, that is, by doing measurements on each requested quantum state and thus getting information about the state. Now, let's back up a bit and clarify these terms. What is a quantum state? A pure state ψ is simply a vector in a Hilbert space. A mixed state, or more simply a state, is a convex combination of pure states ψ_i , that is, a classical probabilistic mixture of pure states:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|, \quad \text{where } \sum_i p_i = 1. \quad (1)$$

Quantum Questions

What is a quantum question? A quantum question is typically called an observable. We'll think of a quantum question as simply an orthonormal basis. The answer to a quantum question will be one of the basis vectors. So, suppose the secret quantum state is a pure state $|\varphi\rangle$ and the quantum question is $\{|\phi_1\rangle, |\phi_2\rangle, \dots, |\phi_M\rangle\}$, a basis of the M -dimensional Hilbert space. According to the basic rules of quantum mechanics, we get the answer $|\phi\rangle$ with probability $|\langle\phi|\varphi\rangle|^2$. If we have a mixed state instead of a pure state, the probability formula is extended by convexity, as usual. How many quantum questions does it take to guess the secret quantum state? That depends on lots of things. It depends on what quantum questions you are allowed to ask me. And it also depends on how different the states in S are from each other. In this context, the word "different" means how distinguishable the states are from each other. For example, two orthogonal pure states are as different as two states can be. Two very nearly parallel pure states are almost indistinguishable in that it takes many experiments and questions to tell them apart based on the outcome statistics. The standard measure of similarity between two pure states is simply their overlap $\langle\phi|\varphi\rangle$. There are measures for the similarity or overlap of mixed states as well, but we won't need the formula. We just need to know that to tell apart two similar states requires many experiments whereas to tell apart two very different states requires few experiments.

Playing search games is much like trying to break codes. If you try to break a code, you want to look for a cryptographic key. To solve classical cryptographic problems with quantum computers, you are looking for a secret key from among a known set of possible secret keys.

So, going back to quantum 20 questions, let's assume you can ask any quantum question you want; that is, you can choose any orthonormal basis as the observable. If all the states in S are sufficiently different from each other, you can find my secret state after only a few questions. Usually, when we use the word "few" in this context, we mean $\log |S|$ or $\log^2 |S|$ or something like that. (A computer scientist would say that few means a polynomial function of the logarithm of the size of the search space.) The key to a fast search is that all the states must be quite different from each other.

It turns out that playing search games is much like trying to break codes. If you try to break a code, you want to look for a cryptographic key. The key is what allows you to decipher the code and read the message. One popular code is the RSA. Named after its inventors—Ronald Rivest, Adi Shamir, and Leonard Adleman—the RSA uses as its key the secret factors of a large number N . Now, suppose you are trying to break a code by finding a secret key k from among a very large set of possible keys $K = \{k_1, k_2, \dots, k_M\}$. Further suppose that, by some process and without knowing the key, you can prepare a quantum state ρ corresponding to the key k . So, you now have a state ρ , which you know comes from the search space $S = \{\rho_1, \dots, \rho_M\}$, which is the set of states corresponding to all the possible secret keys, but you don't know exactly which of the states you have. If the states of S are all sufficiently different, then you can ask quantum questions to determine the secret state efficiently. And if you can find the secret state, then you can easily figure out the original secret key corresponding to that secret state!

Indeed, this is precisely how quantum computers would solve various classical cryptographic problems, such as factoring and finding discrete logarithms. A factoring problem is one in which you are given a very large number N (say, one with 2000 digits), which is the product of two primes $N = pq$, and your task is to find p and q . For the discrete logarithm problem, you are given a large prime number p (say, once again, one with 2000 digits) and two numbers a and b less than p . Your task is to find n such that $a^n = b \pmod{p}$. In both cases, you are looking for a secret key k from among a known set of possible secret keys. Also, in both cases there is a process by which you can prepare a quantum state from which k can be deduced. Significantly, this preparation process does not require knowing k .

This last point is important because, if you had to know the key first, then the code-breaking machine would not be very useful. We will later illustrate this process in an

example (see the section “Simon’s Problem”). Finally, this process has the special and important quality that, for two different keys, k_1 and k_2 , the resulting quantum states, ρ_1 and ρ_2 , are quite different, or clearly distinguishable from one another, as discussed before. We can therefore ask quantum questions, which allow us to distinguish among states and identify secret keys. This ability to distinguish among the states is usually accomplished by eliminating the possibility of a constant fraction, say $1/2$, of the remaining states. As we saw in the game of 20 questions, eliminating a constant fraction after each question allows us to narrow the possible states down to the one true state in only $\log N$ questions. However, since the quantum formula gives probabilities for certain outcomes, we eliminate the false states with high probability (not with certainty), as in the game of random 20 questions.

Identifying Secret Quantum States

Let us fill in some of the technical details of our sketch. First, can we really ask any quantum question? No, we can’t, but fortunately we are able to ask the questions that let us solve factoring and discrete logarithm problems. Recalling our observation that a computation is actually a physical process, we must be sure to carry out efficiently the physical process corresponding to the quantum question we wish to ask. We accomplish this task by breaking down the observable into elementary quantum “gates.” Elementary quantum gates are analogous to the basic logical gates **and**, **or**, and **not**, which are the building blocks of circuits in classical computers (for more details, see Shor 1997). In the case of factoring and discrete logarithm problems, it turns out that we have to ask only one quantum question over and over again in order to obtain enough information for identifying the secret quantum state. Called the quantum Fourier transform, this quantum question allows us to distinguish among the states that arise in the two search spaces for the factoring and discrete logarithm problems. These states are called hidden subgroup states because, in those problems, the key we are looking for corresponds to an unknown subgroup H of a finite abelian group G . The search space corresponds to the set $\{\rho_{H_1}, \rho_{H_2}, \dots, \rho_{H_i}\}$, where H_1 to H_i is a range over all the possible subgroups of G , and ρ_H is the mixed state that corresponds to a uniform mixture of the pure coset states

In factoring and discrete logarithm problems, we must ask only one quantum question over and over again in order to obtain information for identifying the secret quantum state.

$$|c + H\rangle = \frac{1}{\sqrt{|H|}} \sum_{h \in H} |c + h\rangle . \quad (2)$$

It can be shown that for H_1 and H_2 , two different subgroups, the corresponding states ρ_{H_1} and ρ_{H_2} are sufficiently different. Mathematically speaking, the overlap of ρ_{H_1} and ρ_{H_2} is less than $1/2$ (Ettinger et al. 1999). For a discussion of the hidden subgroup problem and the reasons why the quantum Fourier transform is the right quantum question, see Ettinger and Peter Hoyer (1999).

Simon’s Problem

To illustrate everything we have discussed, let’s consider a concrete example known as Simon’s problem. Simon’s problem and the quantum algorithm to solve it contain the essence of what is going on in the factoring and discrete logarithm problems; the latter set of problems, however, also contains a number of technical twists that obscure the main ideas. The set of all bit strings of length n , denoted Z_2^n , is a commutative group if

Our quantum algorithm for solving Simon's problem allows distinguishing among different states and thus discovering the underlying secret bit string.

we add bit strings using “binary add without carry.” This group will be our search space. I will secretly choose an element s of this group and provide you with a function in the form of a “black box,” f_s on Z_2^n , with the following special property: I guarantee that $f_s(x) = f_s(y)$ if and only if $x - y = s$. So, the function f_s encodes the secret bit string s . Because f depends entirely on s , the latter becomes a subscript on f . If you compute the function on the elements of the group $f_s(a), f_s(b), f_s(c), \dots$, eventually you’ll get a collision, which means that you’ll find $f_s(g) = f_s(t)$ and then you’ll know that the secret bit string is $s = g - t$. But notice that the search space, or the group, has 2^n elements, which is a very large number. In the worst case, it could take you $2^{n-1} + 1$ calculations to get a collision, and on average it will take about $2^{n/2}$ because of the so-called birthday paradox.¹ That is still a lot of time! But the quantum algorithm can solve this problem much more quickly—in about n tries only.

Here is how Simon’s problem works: You start with a quantum computer whose qubits are conceptually divided into two registers. Then you prepare the pure state $|\psi\rangle = 1/2^{n/2} \sum_b |b\rangle$, where $b \in Z_2^n$. Thus, in the first register, there is a superposition of all the b s. Now, because you have the black box function f_s , you can compute $f_s(b)$ in the second register to obtain the pure state $|\psi_s\rangle = 1/2^{n/2} \sum_b |b\rangle |f_s(b)\rangle$, where again $b \in Z_2^n$. Notice that this procedure for preparing the state ψ_s is easily accomplished without any knowledge of the secret bit string s . Of course, for different secret bit strings, we obtain different states. In fact, this is the key point. Our quantum algorithm is really just a method used to distinguish among these different states and thus discover the underlying secret bit string.

We now observe, or perform a measurement, on the second register. Because of the way quantum mechanics works, this observation collapses $|\psi_s\rangle$, producing a specific value in the second register, say c , and the first register is left in a superposition of bit strings that map to c under f_s . Because f_s has the special property described earlier, the bit strings that map to c will differ by the secret bit string s . Therefore, the state of the computer is

$$|\psi\rangle_{a,s} = \frac{1}{\sqrt{2}} |a\rangle |c\rangle + \frac{1}{\sqrt{2}} |a + s\rangle |c\rangle, \tag{3}$$

where a and $a + s$ are elements of Z_2^n such that $f_s(a) = c$ and $f_s(a + s) = c$. The only use of the second register is to produce this special superposition in the first register. We will no longer use the second register or its contents, so we drop it from our notation and write

$$|\psi\rangle_{a,s} = \frac{1}{\sqrt{2}} |a\rangle + \frac{1}{\sqrt{2}} |a + s\rangle. \tag{4}$$

When c is chosen, the resulting mixed state can be written as

$$\rho_s = \frac{1}{2^{n-1}} \sum_a |\psi\rangle_{a,s} \langle\psi|_{a,s}. \tag{5}$$

Recall that we don’t know the secret bit string s , and therefore we don’t know that the state we just prepared is ρ_s . All we know is that we have prepared a state that is in the search space of quantum states $\{\rho_s\}_{s \in Z_2^n}$. Each of these possible quantum states corresponds to a possible secret bit string. Our task is to identify the secret quantum state

¹ The birthday paradox derives its name from the surprising result that you only need 23 people (a slightly larger number than $365^{1/2}$) to have a 50 percent chance that at least two of them have the same birthday.

and thus the secret bit string. We now define the Fourier observable. For each bit string b in Z_2^n , define

$$|\chi_b\rangle = \frac{1}{\sqrt{2^n}} \sum_{d \in Z_2^n} (-1)^{b \cdot d} |d\rangle, \quad \text{where } b \cdot d = \sum_i b_i d_i \pmod{2}. \quad (6)$$

The orthonormal basis is $\{|\chi_b\rangle\}$, where $b \in Z_2^n$ is called the Fourier basis or the Fourier observable. Mathematicians might recognize this basis as being composed of the characters of the group Z_2^n . A character χ of a finite abelian group is a homomorphism from the group to the circle in the complex plane. Formally, the Hilbert space in which we are working is $C[G]$, the group algebra, which is the complex vector space with the canonical basis, or the point mass basis, indexed by the elements of the group. A character can be viewed as a vector in $C[G]$ via the following identification:

$$|\chi\rangle = \frac{1}{\sqrt{|G|}} \sum_{g \in G} \chi(g) |g\rangle. \quad (7)$$

It is a fundamental fact (Tolimieri et al. 1997) that the set of characters viewed as vectors in this way is an orthonormal basis for $C[G]$. Indeed, a Fourier transform is nothing other than a change of basis from the point mass basis, $\{|g\rangle\}_{g \in G}$, to the basis of characters, $\{|\chi\rangle\}_\chi$.

It is easy to show (Jozsa 1998) that, if we now observe the contents of the remaining register in the Fourier basis, we observe $|\chi_b\rangle$ with nonzero probability if and only if $s \cdot b = 0 \pmod{2}$. This is the important relationship between the secret bit string s and the only possible outcomes of the experiment. Therefore, if the actual outcome of the observation is $|\chi_b\rangle$, then we have eliminated half of the possible secret states. We have therefore eliminated all states ρ_d such that $d \cdot b = 1 \pmod{2}$. By repeating the state preparation procedure followed by a measurement in the Fourier basis approximately n times, we eliminate all possible states except the true secret state ρ_s . ■

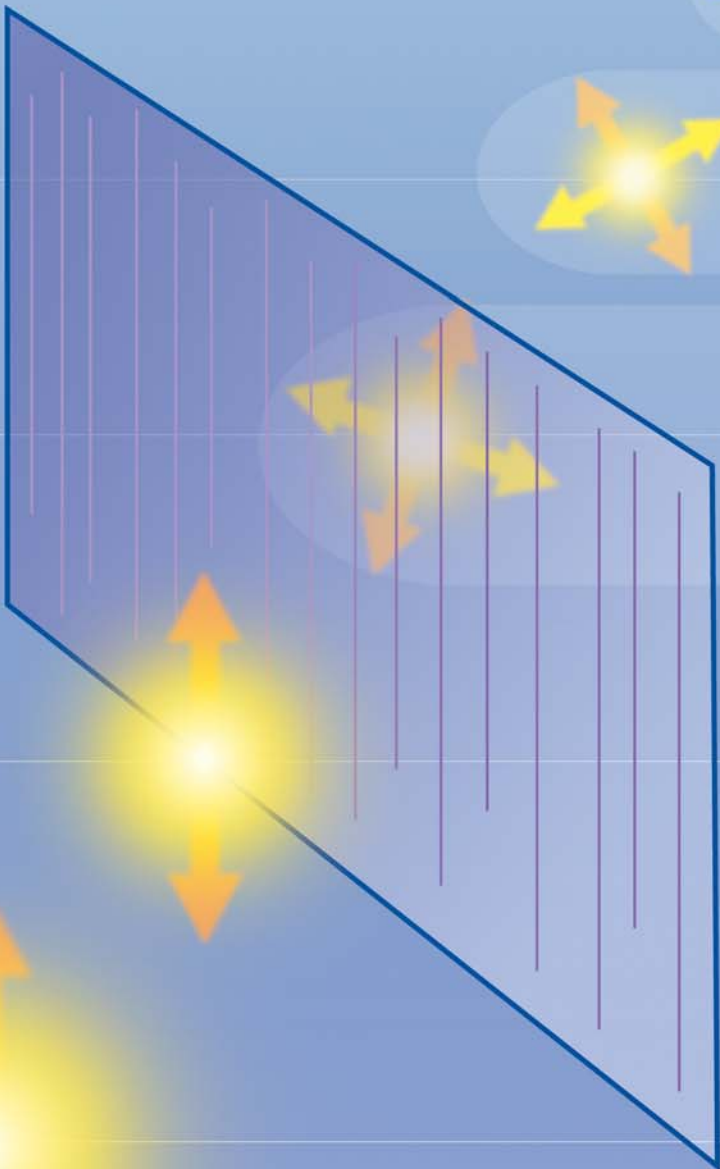
Further Reading

- Ettinger, M., and P. Hoyer. 1999. Quantum State Detection via Elimination [Online]: <http://eprints.lanl.gov/quant-ph/9905099>.
- Ettinger, M., P. Hoyer, and E. Knill. 1999. Hidden Subgroup States are Almost Orthogonal. [Online]: <http://eprints.lanl.gov/quant-ph/9901034>.
- Jozsa, R. 1998. Quantum Algorithms and the Fourier Transform. *Proc. R. Soc. London, Ser. A* **454**: 323.
- Shor, P. W. 1997. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on Logarithms on a Quantum Computer. *SIAM J. Computing* **26**: 1484.
- Tolimieri, R., M. An, and C. Lu. 1997. *Mathematics of Multidimensional Fourier Transform Algorithms*. New York: Springer.

Mark Ettinger graduated from the Massachusetts Institute of Technology in 1987 with Bachelor's degrees in physics and mathematics. In 1996, Mark received his Ph.D. in mathematics from the University of Wisconsin at Madison.



He first came to Los Alamos National Laboratory in 1993, as a graduate student, then entered the postdoctoral program after graduation in 1996, and became a staff member in 1999. Mark worked on the group-theoretical approach to quantum algorithms for four years and is now primarily interested in (classical) algorithmic problems in postgenomic computational biology.



“When two systems, of which we know the states by their respective representatives, enter into temporary physical interaction due to known forces between them, and when after a time of mutual influence the systems separate again, then they can no longer be described in the same way as before, viz. by endowing each of them with a representative of its own. I would not call that one but rather the characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought. By the interaction, the two representatives (or ψ -functions) have become entangled.”

—Erwin Schrödinger (1935)

Quantum State Entanglement

Creation, characterization, and application

Daniel F. V. James and Paul G. Kwiat

Entanglement, a strong and inherently nonclassical correlation between two or more distinct physical systems, was described by Erwin Schrödinger, a pioneer of quantum theory, as “the characteristic trait of quantum mechanics.” For many years, entangled states were relegated to being the subject of philosophical arguments or were used only in experiments aimed at investigating the fundamental foundations of physics. In the past decade, however, entangled states have become a central resource in the emerging field of quantum information science, which can be roughly defined as the application of quantum physics phenomena to the storage, communication, and processing of information.

The direct application of entangled states to quantum-based technologies, such as quantum state teleportation or quantum cryptography, is being investigated at Los Alamos National Laboratory, as well as other institutions in the United States and abroad. These new technologies offer exciting prospects for commercial applications and may have important national-security implications. Furthermore, entanglement is a sine qua non for the more ambi-

tious technological goal of practical quantum computation.

In this article, we will describe what entanglement is, how we have created entangled quantum states of photon pairs, how entanglement can be measured, and some of its applications to quantum technologies.

Classical Correlation and Quantum State Entanglement

To describe the concept of quantum entanglement, we are first going to describe what it is not! Let us imagine the simple experiment illustrated in Figure 1. In that experiment, a source S_1 continually emits pairs of photons in two directions. As seen in the figure, one photon goes toward an observer named Alice, while the other goes toward Bob.

First, imagine that the photons emitted by S_1 are always polarized in the horizontal direction. Mathematically, we say that each photon is in the pure state denoted by the ket $|H\rangle$, that is, the “representative” of the state Schrödinger referred to in the quotation on the opposite page. Because the photons are paired, the combined state of the

two photons is denoted $|HH\rangle$, where the first letter refers to Alice’s photon and the second to Bob’s.

Alice and Bob want to measure the polarization state of their respective photons. To do so, each uses a rotatable, linear polarizer, a device that has an intrinsic transmission axis for photons. For a given angle ϕ between the photon’s polarization vector and the polarizer’s transmission axis, the photon will be transmitted with a probability equal to $\cos^2\phi$. (See the box “Photons, Polarizers, and Projection” on page 76.) Formally, the polarizer acts like a quantum-mechanical projection operator P_ϕ selecting out the component of the photon wave function that lines up with the transmission axis. We say that the polarizer “collapses” the photon wave function to a definite state of polarization. If, for example, the polarizer is set to an angle θ with respect to the horizontal, then a horizontally polarized photon is either projected into the state $|\theta\rangle$ with probability $\cos^2\theta$ or absorbed with probability $1 - \cos^2\theta = \sin^2\theta$. The bizarre aspect of quantum mechanics is that the projection process is probabilistic. The fate of any given photon is completely

This article is dedicated to the memory of Professor Leonard Mandel, one of the pioneers of experimental quantum optics, whose profound scientific insights and gentlemanly bearing will be sorely missed.

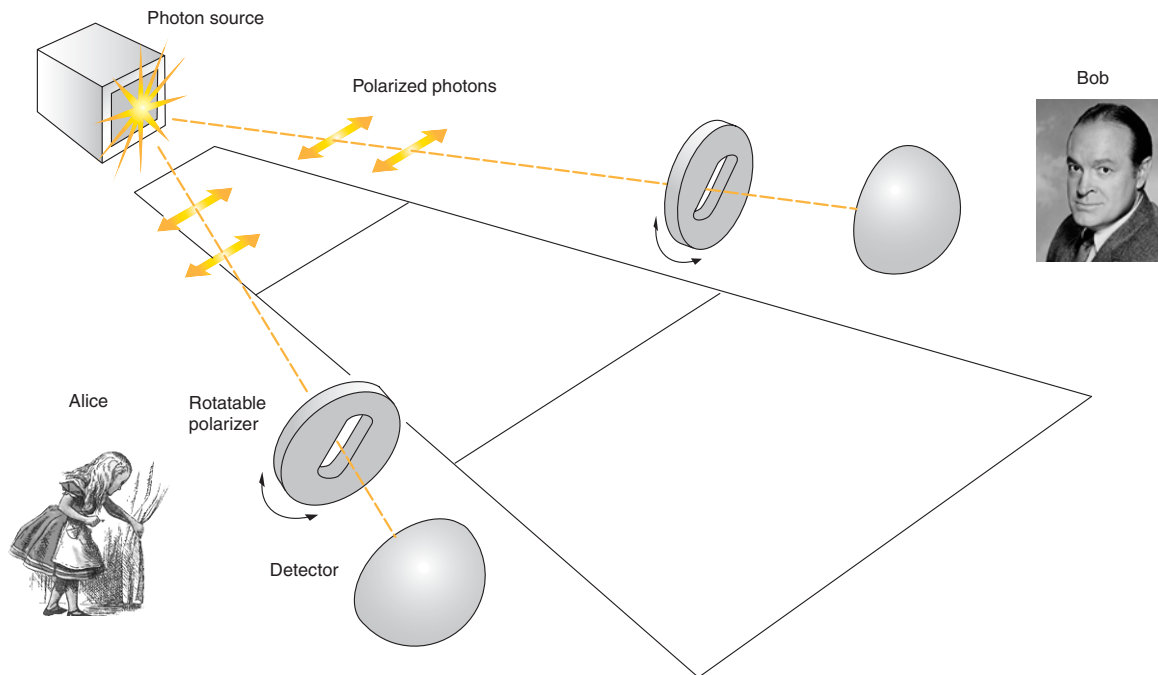


Figure 1. A Simple Two-Photon Correlation Experiment

In this experiment, a source emits pairs of photons: One photon is going to Alice and the other to Bob. Each photon passes through a linear polarizer on its way to its respective detector. Both Alice and Bob’s polarizers are rotatable and can be aligned to any angle with respect to the horizontal, but Bob’s is always kept parallel to Alice’s. For a given polarizer setting, Alice and Bob record those instances when they have the same results, that is, when both detect photons or when they don’t. The figure shows the source emitting two horizontal photons in the state $|\Psi\rangle = |HH\rangle$. The experiment can be performed with other sources to examine differences between other two-photon states. (Picture of Bob is courtesy of Hope Enterprises, Inc.)

unknown. Furthermore, any information about the photon’s previous polarization state is lost.

Getting back to the experiment, we assume that Alice and Bob’s polarizers are always aligned in the same way: When Alice sets her polarizer to a certain angle, she communicates her choice to Bob, who uses the same setting. Behind each polarizer is a detector. In our experiment, Alice and Bob rotate their polarizers to a certain angle with respect to the horizontal and record whether they detect a photon. Then, they repeat the procedure for different polarizer settings. If Alice looks only at her own data (or Bob looks only at his), she can determine the polarization state of the photons emitted by the source—see Figure 2(a). But Alice and Bob can also make a photon-per-photon comparison of their data and determine the probability that they have

the same result, that is, they can examine the photon–photon correlations.

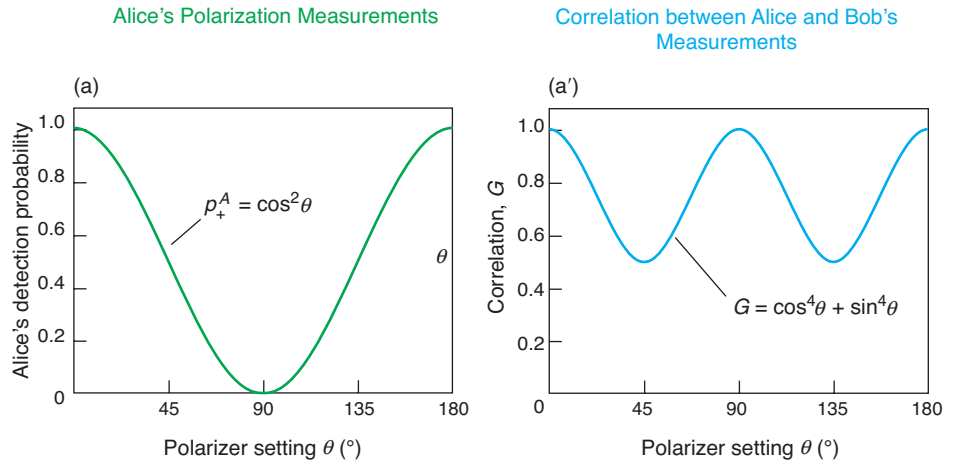
Suppose Alice has her polarizer oriented to transmit horizontally polarized photons. In that case, each photon coming to her from S_1 will be transmitted, and her detector will “click,” indicating a photon has arrived. Subsequent communication with Bob would reveal that he also detected each photon, so at this polarizer setting, there is a perfect correlation between Alice’s detection of a photon and Bob’s. Similarly, by rotating the polarizer to the vertical position, the two would again discover a perfect correlation, namely, neither party would detect his or her photons.

The correlation changes when Alice and Bob have their polarizers oriented, say, at $+45^\circ$ to the horizontal. In that case, the photon sent to Alice has a 50 percent chance of passing through

her polarizer, and independently, the photon sent to Bob has a 50 percent chance of passing through his. The probability is therefore 25 percent that both Alice and Bob detect a photon, 25 percent that neither detects a photon, and thus 50 percent that they obtain the same result.

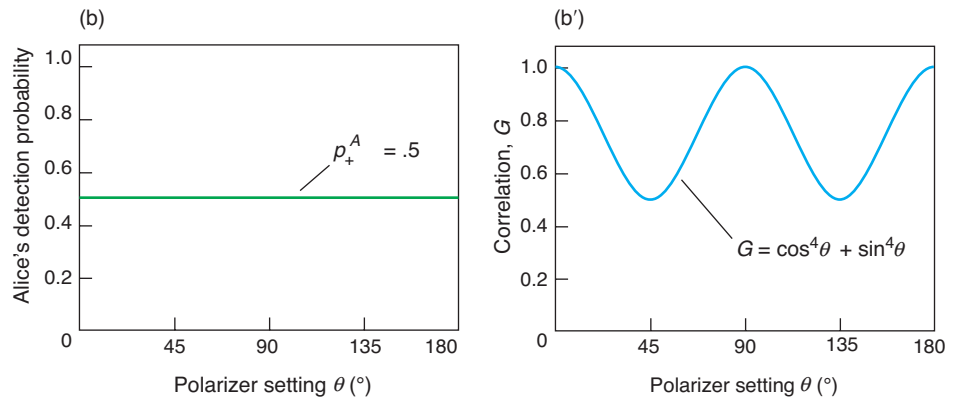
The correlation function G is shown in Figure 2(a’). It is equal to the product of the independent probabilities for detecting a photon $[(\cos^2\theta)_A \times (\cos^2\theta)_B]$, plus the product of the probabilities for not detecting one $[(\sin^2\theta)_A \times (\sin^2\theta)_B]$, where subscripts A and B are for Alice and Bob, respectively. Thus, Alice and Bob deduce that the two photons are independent of each other and the wave function is in fact separable: $|HH\rangle = |H\rangle|H\rangle$. In other words, the correlation is entirely consistent with classical probability theory. The photons are classically correlated.

(a) S_1 emits photons in the pure state $|HH\rangle$. Alice measures a $\cos^2\theta$ function for her polarization data and deduces that photons coming to her are horizontally polarized. (A different linear polarization would shift the curve to the left or right.) (a') We define the correlation function G as the probability that both Alice and Bob detect a photon, plus the probability that neither detects a photon. For this source, G is completely consistent with classical probability theory for independent events; that is, the correlation function is the product of the detection probability of each photon in the pair.



Probability that Alice (or Bob) detects a photon: $p_+ = \cos^2\theta$.
 Probability that Alice (or Bob) does not detect a photon: $p_- = \sin^2\theta$.
 For independent photons: $G = G_{HH} = p_+^A \times p_+^B + p_-^A \times p_-^B = \cos^4\theta + \sin^4\theta$.

(b) The source S_2 emits photons in the partially mixed state $1/2(|HH\rangle\langle HH| + |VV\rangle\langle VV|)$. Photons from this source do not have a net polarization. Alice receives at random either an $|H\rangle$ or a $|V\rangle$ photon, so the sum of her measurements averages to a 50 percent detection probability independent of angle. (b') The correlation function G , however, is the same as in (a), revealing that the photons in each pair are independent of each other and have polarization H or V. Therefore, the two photons exhibit the same classical correlations seen in (a).



For this mixed state,
 $G = 1/2(G_{HH} + G_{VV}) = G_{HH}$.

(c) The source S_3 emits photons in the maximally entangled state $1/\sqrt{2}(|HH\rangle + |VV\rangle)$. Unlike the photons in the mixed state, each photon is unpolarized. Nevertheless, if Alice and Bob align their polarizers the same way, they always get the same result independent of angle. (c') Polarization measurements of the two photons are 100 percent correlated. The photons exhibit "quantum" correlations.

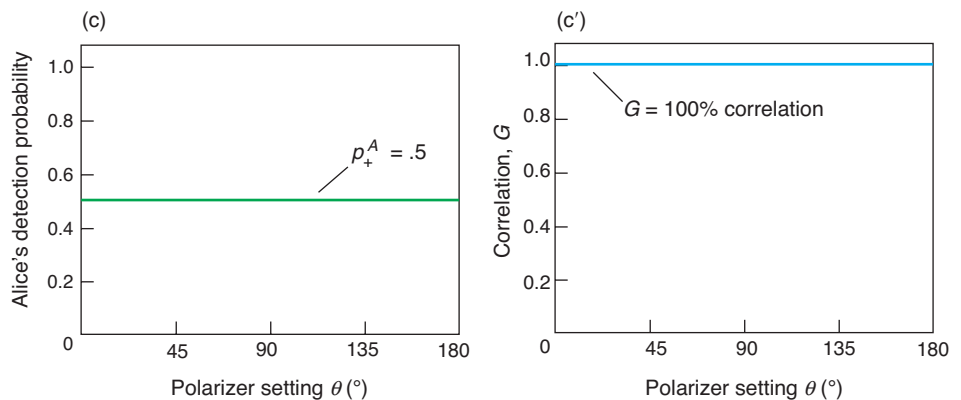


Figure 2. Quantum States, Polarization, and Correlation

The three sets of graphs show the results of the three experiments discussed in the text. In each case, the leftmost graph shows the probability that Alice alone detects a photon and reveals information about the net polarization state of her photon. The rightmost graph shows the probability that Alice and Bob have the same result, which reveals information about the two-photon state.

Pure, Entangled, or Mixed?

A pure state is a vector in a system’s Hilbert space. For example, the most general, pure two-photon polarization state can be written as

$$|\psi_{\text{pure}}\rangle = \alpha|HH\rangle + \beta|HV\rangle + \gamma|VH\rangle + \delta|VV\rangle . \quad (1)$$

This state is specified by the four probability amplitudes α , β , γ , and δ (expressed by four complex numbers or eight real numbers) although these parameters are subject to two constraints. The first is that the mean-square amplitudes must equal unity, that is,

$$|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1 . \quad (2)$$

The second relates to the fact that the overall phase of a wave function has no physical relevance. The net result of these constraints is that any pure two-photon state depends on only six independent real numbers.

In general, however, any physical system contains a greater or lesser degree of randomness and disorder, and one must adapt the formalism of quantum mechanics to take this randomness into account. We do so by averaging over the fluctuations. It is convenient to represent states as density operators, or density matrices, formally defined as

$$\rho = \overline{|\psi\rangle\langle\psi|} , \quad (3)$$

where the overbar denotes an ensemble average over the randomness. All the measurable properties of the state are determined by ρ .

The density matrix must be used when representing mixed states, which can be thought of as probabilistic combina-

tions of pure states. Mathematically, the density matrix can always be decomposed into an incoherent sum over pure states,

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| , \quad (4)$$

where each $|\psi_i\rangle$ is a pure state and p_i are probabilities with values that lie between 0 and 1 and whose sum is 1. In general, this decomposition is not unique. To characterize mixed states, one uses mean values and classical coherences; that is, one must specify the four mean-square amplitudes (subject to the constraint $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$) and the six independent classical complex correlations $\overline{\alpha^*\beta}$, $\overline{\alpha^*\gamma}$, and so on.

For example, the source S_2 mentioned in the text emits a partially mixed state that is 50 percent $|HH\rangle$ and 50 percent $|VV\rangle$, so that

$$\rho_{\text{mix}} = 0.5 |HH\rangle\langle HH| + 0.5 |VV\rangle\langle VV| , \quad (5)$$

or in matrix form

$$\rho_{\text{mix}} = \begin{bmatrix} .5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .5 \end{bmatrix} . \quad (6)$$

This state is neither pure nor completely random; it is partially mixed.

We next consider whether quantum states involving two or more systems (for example, two photons), are separable

Now, consider performing the experiment with a second source S_2 that has a 50 percent chance to emit two horizontally polarized photons $|HH\rangle$ and a 50 percent chance to emit two vertically polarized photons $|VV\rangle$. This type of source emits photons in a mixed state, which cannot be written as a single “ket.” Instead, a mixed state must be analyzed in terms of several kets, each representing a particular, distinct pure state that has a probability associated with it. Making

a measurement on a mixed state is equivalent to probing an ensemble of pure states. The likelihood of measuring a particular pure state is given by the appropriate probability. (More-detailed, mathematical descriptions of pure and mixed quantum states are found in the box “Pure, Entangled, or Mixed?” above.)

The output of S_2 is random (either $|HH\rangle$ or $|VV\rangle$), so Alice receives at random either an $|H\rangle$ or a $|V\rangle$ photon. Because the probability of detecting

$|H\rangle$ is $1/2 \cos^2\theta$, and the probability of detecting $|V\rangle$ is $1/2 \sin^2\theta$, Alice has a 50 percent chance of detecting a photon regardless of how she sets her polarizer. The same is true for Bob. Each observer, therefore, deduces that the photons coming from S_2 have no net polarization. But as seen in Figure 2(b’), the correlation function tells a different story. In fact, the correlation function for this source is identical to the one obtained for S_1 because, in both cases, the individual

or entangled. If the state is separable and pure, it can be written (in some basis) as a product of the states of the individual systems, that is, as

$$|\psi\rangle = |\psi_A\rangle \otimes |\psi_B\rangle, \tag{7}$$

where \otimes denotes the tensor product. The state $|\psi_1\rangle = |HH\rangle$ is one such product of pure states and can be written as

$$|\psi_1\rangle = |H_A\rangle \otimes |H_B\rangle. \tag{8}$$

Another example is the state

$$|\psi\rangle = (|HH\rangle + |HV\rangle + |VH\rangle + |VV\rangle)/2, \tag{9}$$

which can be written as the product state

$$|\psi\rangle = 1/\sqrt{2}(|H\rangle + |V\rangle)_A \otimes 1/\sqrt{2}(|H\rangle + |V\rangle)_B. \tag{10}$$

A third example is the matrix ρ_{mix} on the opposite page, which represents a separable mixed state.

In contrast, if there is no way to write the two-photon state as a direct product of states, the state is said to be entangled. This definition leads to a quantity called concurrence, which is defined for the general pure state $|\psi_{\text{pure}}\rangle$ by

$$C = 2|\alpha\delta - \beta\gamma|. \tag{11}$$

If and only if C is zero is the state separable. If C is equal to unity (its maximum value), the state is maximally entangled.

For example, consider any one of the four Bell states

$$|\Phi_{\pm}\rangle = 1/\sqrt{2}(|HH\rangle \pm |VV\rangle), \text{ and}$$

$$|\Psi_{\pm}\rangle = 1/\sqrt{2}(|HV\rangle \pm |VH\rangle). \tag{12}$$

These states are a basis for the two-photon Hilbert space, and linear combinations of the four states can be used to represent any two-photon state. If we compare, say, $|\Phi_{+}\rangle$ with the general state $|\psi_{\text{pure}}\rangle$, we have $\alpha = \delta = 1/\sqrt{2}$, and $\beta = \gamma = 0$. Thus $C = 1$, and this Bell state is maximally entangled (as are the other three).

The value of C provides a good metric for the amount of entanglement in a pure two-qubit system. Equivalently, some researchers use C^2 (a quantity known as the tangle) to characterize the degree of entanglement.

The concurrence can also be defined for mixed states, although the definition is much more complicated. Indeed, calculating the concurrence for mixed states of more than two qubits is currently a hot topic of research.

In the everyday world, it is common to ascribe two (or more) variables to the same object (for example, a hot, sweet cup of coffee). Similarly, quantum states are described by the two characteristics discussed above, so that it is possible to have a pure entangled state, a pure separable state, a mixed separable state, or something in between, such as a partially mixed, partially entangled state.

photons leave the source in definite polarization states. For S_1 , the polarization information is “carried” individually by each photon. For S_2 , the polarization information is carried by the photon pairs. By examining the correlations, Alice and Bob can deduce that information.

A different situation occurs for a source S_3 that emits pairs of photons in the state $|\Phi_{+}\rangle = 1/\sqrt{2}(|HH\rangle + |VV\rangle)$. Like the mixed state from S_2 , this state is a combination of two horizon-

tally polarized photons and two vertically polarized photons. Unlike the mixed state, $|\Phi_{+}\rangle$ is a coherent, quantum mechanical superposition: A probability amplitude is associated with each component, $|HH\rangle$ and $|VV\rangle$, and the two components have a fixed phase relationship. An important property of this particular state is that we can rotate the axes of polarization (H and V) and not change the state’s essential properties.

The state $|\Phi_{+}\rangle$ is a fully entangled

quantum state. It cannot be factorized, or separated, into a part describing one of the photons and a part describing the other. The two photons are inextricably linked to each other and their properties are always correlated. A measurement of one of the photons makes the two-photon state instantly disappear, and the remaining photon assumes a definite state that is perfectly correlated with the measured photon. Neither photon carries definite information by itself—all the information is

carried in the joint two-photon state.

Thus, when Alice and Bob repeat the experiment using the source S_3 , the correlation is 100 percent regardless of polarizer orientation (assuming Bob's polarizer is always set the same way as Alice's). Figure 2(c) illustrates the striking difference between the classical correlations of the photons generated by the sources S_1 and S_2 and the nonclassical correlations exhibited by entangled photons.

To better understand the correlation curve shown for $|\Phi_+\rangle$, consider that quantum mechanics allows us to express that state in any basis; that is, $|\Phi_+\rangle = 1/\sqrt{2} (|XX\rangle + |YY\rangle)$, where $|X\rangle$ is an arbitrary linear-basis state and $|Y\rangle$ is the orthogonal-basis state. Suppose Alice has her polarizer set to $+45^\circ$. In the diagonal ($+45/-45$) basis, the entangled state will be $|\Phi_+\rangle = 1/\sqrt{2} (|+45,+45\rangle + |-45,-45\rangle)$. If Alice detects her photon (a 50–50 proposition), then Bob's photon will collapse to the $|+45\rangle$ state, and he will detect his photon as well. Likewise, if Alice doesn't detect her photon, Bob won't detect his. The same deductions can be made for any polarizer setting.

According to quantum mechanics, the correlation occurs regardless of the distance separating the two photons. For example, suppose one of two entangled photons from the state $|\Phi_+\rangle$ is sent to Alice, who "stores" it in her laboratory at Los Alamos, New Mexico. The other photon is sent to Bob, who is in orbit about the star α -Centauri, nearly 4 light-years away. After some time, Alice performs a measurement on her photon and determines that it is $|H\rangle$. Her measurement selects the $|HH\rangle$ part of the state $|\Phi_+\rangle$ and eliminates the $|VV\rangle$ part so that Bob's photon is necessarily in the state $|H\rangle$. If, instead, Alice has determined that her photon was $|+45\rangle$, the state of Bob's photon will be instantly collapsed to $|+45\rangle$ as well. In other words, the state of Bob's photon

has been nonlocally influenced by Alice's measurement. By nonlocal, we mean that the correlation between Alice and Bob's measurements occurs even if there is not enough time for a light signal (or any signal) to propagate between the two experimentalists. This is not to say that special relativity has been violated: Because Alice cannot predetermine the outcome of her measurement, she cannot use the nonlocal quantum correlations to send any information to Bob. In fact, entanglement can never be used to send signals faster than the speed of light. Nonetheless, Bob's photon "knows" the outcome of Alice's measurement.

Nonlocality was the central point of a famous argument raised by Albert Einstein, Boris Podolsky, and Nathan Rosen in 1935, now known as the EPR paradox. The three physicists disagreed with the Copenhagen interpretation of quantum mechanics, according to which the state of a quantum system is indeterminate until it is projected into a definite state as a result of a measurement. Einstein, Podolsky, and Rosen argued that even unmeasured quantities corresponded to definite "elements of reality." The quantum state only appeared to be indeterminate because some of the parameters that characterize the system were unknown and unmeasurable. These local parameters, or "hidden variables," determined the outcome of the experiment.

In 1964, John Bell showed that the correlations between measured properties of any classical two-particle system would obey a mathematical inequality, but the same measured correlations would violate the inequality if the two particles were an entangled quantum system. Experiments could

therefore determine if nature exhibited nonlocal features. Following the development of laboratory sources of entangled photons, experimental tests of Bell's inequality were pursued with vigor. The results to date suggest that the observed photon correlations cannot be explained by any local hidden-variable theory,¹ and most physicists agree that quantum mechanics is truly a nonlocal theory.

Entanglement and Quantum Information

Entanglement, a measurable property of quantum systems, can be exploited for specific goals. Here, we present three potential applications, all of which have been shown to work as proof-of-principle demonstrations in the laboratory.

Quantum Cryptography. Consider two bank managers, Alice and Bob, who want to have a secret conversation. If they are together in the same room, they can simply whisper discretely to each other, but when Alice and Bob are in their respective cross-town offices, their best chance for secret communication is to encrypt their messages.

A generic, classical encryption protocol would begin when Alice and Bob convert their messages to separate binary streams of 0s and 1s. Encryption (locking up the messages) and decryption (unlocking the messages) are then performed with a set of secret "keys" known only to the two bankers. Each key is a random string of 0s and 1s that is as long as the binary string comprising each

¹ There were two loopholes to the EPR tests. The first stemmed from the fact that the detectors were not efficient enough. Consequently, the observed correlations could have been the result of some new physics that did not require nonlocal interactions. The second loophole stemmed from the researchers' inability to choose rapidly and randomly a basis for photon measurement. This inability allowed for a potential communication conspiracy between Alice and Bob's systems. Both of these loopholes have recently been closed but, so far, not in the same experiment.

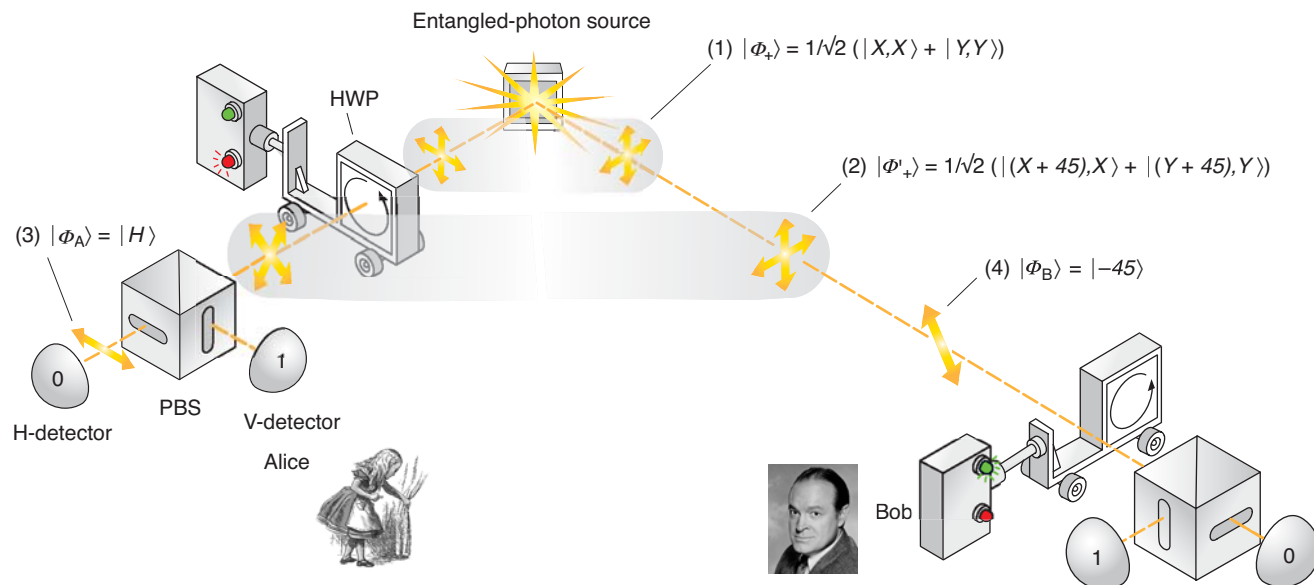


Figure 3. Quantum Cryptography Using Entangled Photons

Alice and Bob can use the properties of entangled photons to create a pair of identical cryptographic keys. (1) The source emits entangled photons in a maximally entangled state $|\Phi\rangle = 1/\sqrt{2}(|XX\rangle + |YY\rangle)$, where $|X\rangle$ is an arbitrary linear-basis state and $|Y\rangle = |X + 90^\circ\rangle$ is the orthogonal-basis state. One photon goes to Alice and the other to Bob. (2) Alice chooses at random either to let her photon pass or to insert a half-wave plate (HWP), which will rotate her photon by $+45^\circ$. The latter choice changes the relative orientation between the two photons by $+45^\circ$. In the case shown, she chooses to rotate her photon. The new entangled state is $|\Phi'\rangle$. (3) Alice uses a polarizing

beam splitter (PBS) to measure her photon in the H/V basis. This optical element transmits horizontally polarized photons and reflects vertically polarized photons, and her unpolarized photon can collapse to either a horizontal or vertical polarization with equal probability. In this case, it collapses to a horizontal polarization. Alice records a bit value of 0. (5) Bob's photon was entangled with Alice's, so as a result of her measurement, his photon assumed the definite polarization state $|H - 45^\circ\rangle = |-45^\circ\rangle$. If Bob makes the same choice as Alice and inserts his HWP, he will rotate his photon's polarization by $+45^\circ$ and into a horizontal polarization. His photon will register in the H-detector,

and he will record a bit value of 0. If he makes the opposite choice and doesn't rotate his photon, the photon polarized at -45° has an equal probability of going to either detector (bit value either 0 or 1). As seen in Table I on the next page, whenever Bob and Alice make the same choice, they keep the bit because their bit values coincide. If they make opposite choices, they discard the bit since the values are not correlated. Alice and Bob can construct an identical sequence of random bits—a cryptographic key—simply by declaring their sequence of choices. The discussion can be public because the bit values are never revealed.

message. To encrypt, Alice (the sender) sequentially adds each bit of the key to each bit of her message, using modulo 2 addition ($0 + 0 = 0$, $0 + 1 = 1$, $1 + 0 = 1$, and $1 + 1 = 0$). She then sends the encrypted message to Bob, who decrypts it simply by repeating the operation, that is, by performing a sequential, bit-by-bit modulo 2 addition of the key to the message.

This type of encryption protocol, known as a one-time pad, is currently the only provably secure protocol. But the one-time pad is effective only if Alice and Bob never reuse the key, and more obviously, if the key remains

secret. A potential eavesdropper, Eve, cannot be allowed to glean any part of the bit stream that makes up the key. Therein lies a central problem of cryptography: How can secret keys be created and then securely distributed? The nonlocal correlations of entangled photons can play a role in this regard. (One can also exploit the properties of nonentangled photons in cryptographic schemes. See the article "Quantum Cryptography" on page 68.)

In the entangled-state quantum cryptography scheme, Alice and Bob perform an experiment similar to the one described in the first section of

the paper. They use a source S_3 that emits entangled photons in the general state $|\Phi_+\rangle = 1/\sqrt{2}(|XX\rangle + |YY\rangle)$, where $|X\rangle$ is an arbitrary linear-basis state and $|Y\rangle$ is the orthogonal-basis state. One photon goes to Alice and the other to Bob. In this protocol, however, either banker can choose—at random and independent of each other—to use a half-wave plate (HWP) to rotate photon polarization by a set amount. The bankers then detect the photon in the H/V basis using a polarizing beam splitter, which transmits horizontally polarized photons and reflects vertically polarized photons (see Figure 3).

Table I. Constructing a Cryptographic Key with Entangled Photons

First Receiver (Alice)			Polarization to Second Receiver	Second Receiver (Bob)			Communication Results
Angle of Rotation (°)	Detector	Bit Value		Angle of Rotation (°)	Detector	Bit Value	
0	H	0	H	0	H	0	Keep bit
0	V	1	V	0	V	1	Keep bit
0	H	0	H	+45	H or V	0 or 1	Discard bit
0	V	1	V	+45	H or V	0 or 1	Discard bit
+45	H	0	-45°	0	H or V	0 or 1	Discard bit
+45	V	1	+45°	0	H or V	0 or 1	Discard bit
+45	H	0	-45°	+45	H	0	Keep bit
+45	V	1	+45°	+45	V	1	Keep bit

Detection of a horizontally polarized photon is recorded as a 0; of a vertically polarized photon, as a 1.

After a sufficient number of measurements (that number is dictated by the length of the key), Alice and Bob have a public discussion, during which they reveal whether they inserted the HWP before each measurement. At no time do they reveal the actual measurement results. Whenever Alice and Bob make the same choice (50 percent of the time), they know from the properties of entangled photons that they will have completely correlated results. By contrast, if one of them uses the HWP and the other doesn't, they will discard the results because their measurements would be completely uncorrelated (see Table I). Following this public discussion, each banker will be able to privately construct the same random string of 0s and 1s—an ideal key for cryptography.

What about the eavesdropper Eve? She is completely foiled in her attempts to know the secret key. Certainly, she cannot tap the photon line, as she might with conventional, classical communications. A single, indivisible quantum object—namely, a photon—is the conveyor of information in this cryptographic protocol. If Eve steals Bob's photon (a "denial-of-service" attack), the pho-

ton's information never becomes part of the key. Thus, although a wiretap would reduce the rate of the transmission, it would not jeopardize the security of the key.

Eve can try to intercept the photon, measure it, and send another one to Bob. But any measurement Eve would make to determine the photon's polarization state would necessarily perturb the photon and collapse the entangled state. The photon she sends to Bob would therefore be classically correlated with Alice's photon. Consequently, Eve's intervention would necessarily induce additional errors into Bob's key.

This last point is significant. Unlike their theoretical counterparts, the encryption keys created by an actual quantum cryptography system initially have a small fraction of errors, because real equipment is always less than perfect. To make sure their key is secure, Alice and Bob ascribe all errors to Eve and then use this "bit error rate" to estimate the maximum amount of information available to the eavesdropper. They then use a privacy amplification scheme (discussed in the cryptography article on page 68) to reduce Eve's knowledge of the secret key to less than one bit.

But the bit error rate alone can lead

to a false sense of security. If nonentangled photons with a definite polarization are sent to Bob, it is conceivable that some other degree of freedom may also be coupled to the polarization state. For example, if separate lasers are used to produce the two polarization states, the photons from each laser may have slightly different timing characteristics or frequency spectra. Such a difference would in principle allow an eavesdropper to distinguish between photons without disturbing the polarization state and, hence, without affecting the bit error rate.

When the photons are entangled, however, any leakage of information to other degrees of freedom can be shown to automatically manifest itself in the error rate detected by Alice and Bob. In other words, any degree of freedom with which the polarization might be coupled will cause noticeable effects on the polarization correlations. Therefore, using only the detected error rates, one can set an upper limit on the information available to an eavesdropper, even one who is not directly measuring the polarization of the photons, and then use privacy amplification to eliminate that information.

As a last resort, Eve may think of "cloning" Alice's photon. She could measure the clone while allowing the

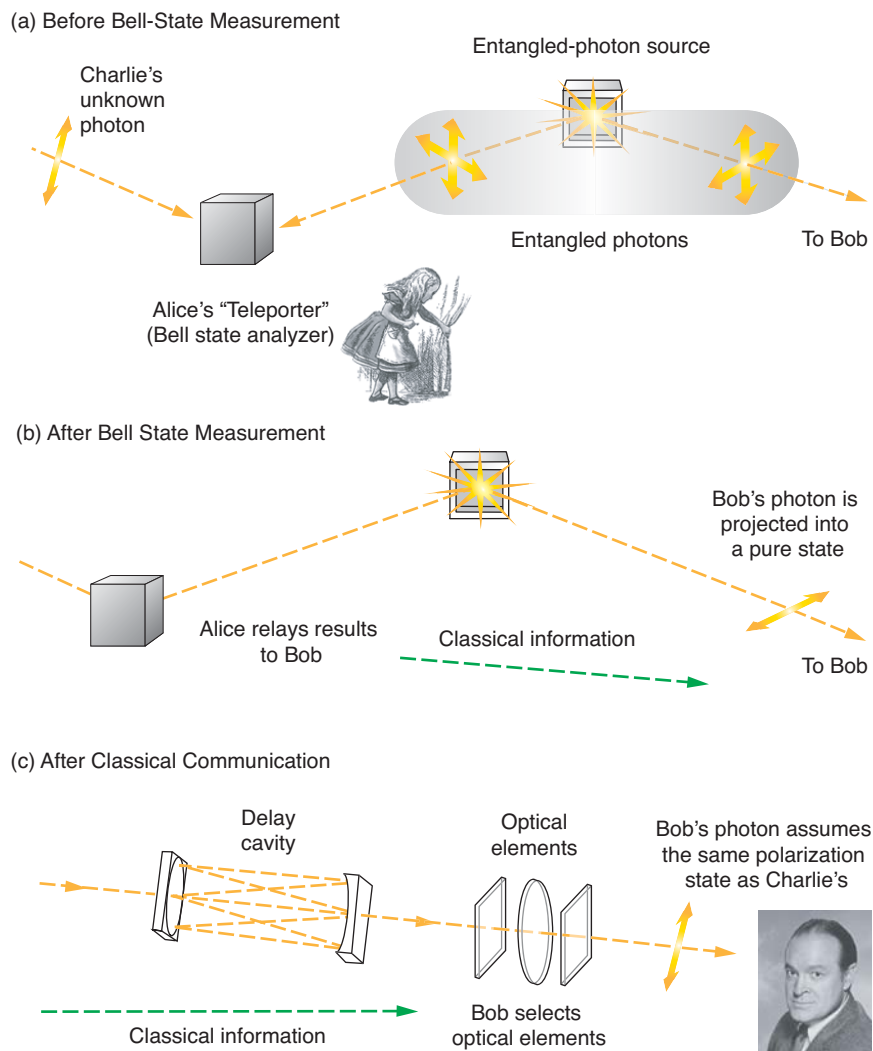


Figure 4. Quantum State Teleportation

(a) Alice's teleportation lab consists of an entangled photon source and a Bell state analyzer (the teleporter). One entangled photon goes to Bob and the other to the teleporter. Charlie sends a photon of unknown polarization state into the teleporter. (b) Alice performs a joint polarization measurement of the two photons in the teleporter and relays the result to Bob using two classical bits of information. The photon going to Bob is projected into a pure state as a result of Alice's measurement. (c) Upon receiving Alice's classical information, Bob performs a simple transformation on his photon, such as a rotation of the polarization vector. He duplicates the polarization state of Charlie's photon without knowing anything about its original state.

original to continue on to Bob, thus completely covering her tracks. But she is again foiled by quantum mechanics. According to the no-cloning theorem, it is impossible to make a copy of a photon in an unknown state while simultaneously preserving the original. (See the box "The No-Cloning Theorem" on page 79.) Eve is clearly out of business.

Teleportation. In 1993, Charles Bennett of IBM, Yorktown Heights, and his colleagues proposed a remarkable experiment with entangled particles, namely, the "teleportation" of a pure quantum state from one location to another.

Charlie wants to send his friend Bob a photon in an arbitrary, pure quantum state $|\psi\rangle = \alpha|H\rangle + \beta|V\rangle$. He enlists the aid of Alice, who happens to run the Teleportation Laboratory shown in Figure 4. Inside the lab, a source S_3 is emitting a pair of entangled photons, one of which goes off to Bob. The other photon is input into Alice's "teleporter." Charlie is instructed to send his photon into the teleporter as well.

Alice then performs a special joint measurement of the polarization state of the two photons in the teleporter. She relays the result to Bob, who subsequently performs a simple transformation of the polarization state of

his photon. As if by magic, the state of Bob's photon is transformed into the state of Charlie's original photon.

Mathematically, this magic is described as follows. The three-photon initial state (that is, Charlie's photon plus the two entangled photons) can be represented as

$$|\psi_0\rangle = (\alpha|H\rangle + \beta|V\rangle)_C \times 1/\sqrt{2}(|HH\rangle + |VV\rangle)_{A,B}, \quad (1)$$

where the subscripts C, A, and B refer to Charlie's, Alice's, and Bob's photons, respectively. But $|\psi_0\rangle$ can also be represented as a superposition of states, each constructed in the following way: Charlie and Alice's photons

are represented by one of the Bell states $|\Phi_{\pm}\rangle = 1/\sqrt{2} (|HH\rangle \pm |VV\rangle)$ and $|\Psi_{\pm}\rangle = 1/\sqrt{2} (|HV\rangle \pm |VH\rangle)$,

and Bob's photon is represented as a photon in a pure state. Thus,

$$|\psi_0\rangle = 1/2\{|\Phi_{-}\rangle_{C,A}(\alpha|H\rangle - \beta|V\rangle)_{B} + |\Phi_{+}\rangle_{C,A}(\alpha|H\rangle + \beta|V\rangle)_{B} + |\Psi_{-}\rangle_{C,A}(-\beta|H\rangle + \alpha|V\rangle)_{B} + |\Psi_{+}\rangle_{C,A}(\beta|H\rangle + \alpha|V\rangle)_{B}\} \quad (2)$$

Technically speaking, this representation is possible because the Bell states are a basis for the two-photon Hilbert space and any state of two photons can be represented as a linear superposition of these states. It is important to point out that Alice's photon remains entangled with Bob's. Teleportation relies on Alice's ability to perform a joint polarization measurement that explicitly projects the two photons in the teleporter into one of the four Bell states. Once Alice completes her measurement, Bob's photon (which is totally correlated to Alice's) will assume the corresponding pure state. For example, if the Bell state measurement produces the result $|\Psi_{-}\rangle_{C,A}$, then Bob's photon would be projected into the pure state $|\psi\rangle = (-\beta|H\rangle + \alpha|V\rangle)_{B}$. By using a simple optical element, Bob can rotate the polarization state of his photon by 90° and transform it into the state $|\psi'\rangle = (\alpha|H\rangle + \beta|V\rangle)_{B}$, that is, the original input state. Provided Alice can specify which Bell state was measured (a specification that requires two bits of classical information), Bob can always choose an appropriate optical element to effect the proper rotation.

In a series of groundbreaking exper-

iments conducted at the University of Innsbruck, Austria, Anton Zeilinger and coworkers were the first to demonstrate quantum teleportation. The group is now able to determine two of the four Bell states unambiguously (the other two states give the same experimental signature²) and prove for those cases that the state of Charlie's photon could indeed be transferred to Bob's.

Several points should be made about quantum teleportation. First, during the entire procedure, neither Alice nor Bob has any idea what the values are for the parameters α and β that specify Charlie's photon. The initial, arbitrary pure state remains unknown. Second, teleportation is not cloning. The original state of Charlie's photon is necessarily destroyed by Alice's measurement, so the photon that Bob ends up with is still one of a kind.

Finally, hopeful sci-fi fans may be a little disappointed by this realization of teleportation. Unlike the TV show "Star Trek," in which Captain Kirk could be transported body and soul from the starship *Enterprise* to the surface of an alien planet,³ here only certain information about the photon is transferred to a photon in some faraway location. Because photons have numerous degrees of freedom in addition to their polarization, the original and the teleported photons are two different entities. And it goes without saying that an even simpler way for Charlie to send his quantum state to Bob would be to dispatch the original photon directly to him.

Nevertheless, teleportation remains an interesting application of quantum state entanglement. Furthermore, researchers have discussed how it might form the basis of a distributed

network of quantum communication channels and how this basic information protocol might be useful for quantum computing.

Quantum Microscopy and Lithography. The general topic of quantum metrology involves capitalizing on the ultrastrong correlations of entangled systems to make measurements more precisely than would be possible with classical tools. The two main photon-based applications under investigation are quantum microscopy and quantum lithography.

At present, two-photon microscopy is widely used to produce high-resolution images, often of biological systems. However, the classical light sources (lasers) used for the imaging have random spreads in the temporal and spatial distributions of the photons, and the light intensity must be very high if two photons are to intersect within a small enough volume and cause a detectable excitation. The high intensity can damage the system under investigation. Because the temporal and spatial correlations may be much stronger between members of an entangled photon pair, one could conceivably get away with much weaker light sources, which would be much less damaging to the systems being observed. Moreover, entangled-photon sources may also enable obtaining enhanced spatial resolution.

Lithography, in which a pattern is optically imaged onto some photoresist material, is the primary method of manufacturing microscale or nanoscale electronic devices. An inherent limitation of this process is that details smaller than a wavelength of light cannot be written reliably. However, quantum state entanglement might circumvent this limitation. Under the right circumstances, the interference pattern formed by beams of entangled photon pairs can have half the classical fringe spacing.

Quantum lithography requires two beams of photons, which we

² Distinguishing between the four Bell states is still an unsolved technical problem. It requires a strong nonlinear interaction between two photons, which is extremely difficult to achieve in practice.

³ "Teleportation" (though it was not explicitly called that) was supposedly introduced in this TV show because the producer, Gene Roddenberry, wished to save the expense of simulating the landing of a starship on a planet.

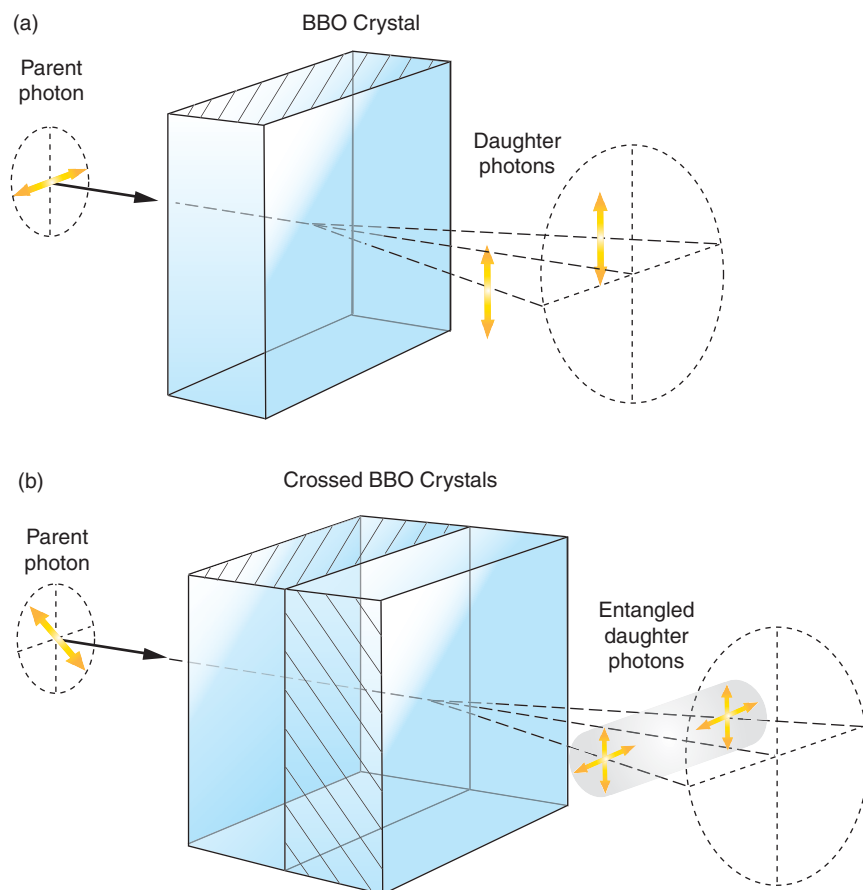


Figure 5. Entangled-Photon Source

(a) For a given orientation of the beta-barium borate (BBO) crystal, a horizontally polarized parent photon produces a pair of vertically polarized daughters. The daughters emerge on opposite sides of an imaginary cone. The cone's axis is parallel to the original direction taken by the parent photon. The two daughter photons are not in an entangled state. Reorienting the BBO crystal by 90° will produce a pair of horizontally polarized daughters if a vertically polarized pump beam is used. (b) Passing a photon polarized at $+45^\circ$ through two crossed BBO crystals can produce two photons in an entangled state. Because of the Heisenberg uncertainty principle, there is no way to tell in which crystal the parent photon "gave birth," and so a coherent superposition of two possible outcomes results: a pair of vertically polarized photons or a pair of horizontally polarized photons. The photons are in the maximally entangled state $|\Phi_+\rangle = 1/\sqrt{2}(|HH\rangle + |VV\rangle)$.

will call A and B, but in this case, the type of entanglement is different from the one discussed in the previous sections. What is needed is a coherent superposition consisting of the state in which two photons are in beam A while none are in B and the state in which no photon is in beam A while

two photons are in B. Such number-entangled states can be made in the laboratory, and the predictions about fringe spacings have been verified. However, other obstacles must be overcome in order to surpass current classical-lithography techniques. Researchers continue to explore the

potential of this idea with the hope of achieving a viable commercial technology.

Creating and Measuring Entangled States

If quantum state entanglement is such a remarkable property because it allows one to perform secret communications, teleport states, or test the nonlocality of quantum mechanics, one naturally wonders how to make entangled states. Currently, scientists can create entangled states of particles in a controlled manner by using several technologies such as ion traps, cavity quantum electrodynamics (QED), and optical down-conversion. Here, we will concentrate on the optical realization.

Crystals of a certain chemical structure, such as beta-barium borate (BBO), have the property of optical nonlinearity, which means that the polarizability of these crystals depends on the square (or higher powers) of an applied electric field. The practical upshot of this property is that, when passing through such a crystal, a single-parent photon can split (or down-convert) into a pair of daughter photons. The probability that this event occurs is extremely small; on average, it happens to only one out of every 10 billion photons!

When down-conversion does occur, energy and momentum are conserved (as they must be for an isolated system). The daughter photons have lower frequencies (longer wavelengths) than the parent photon and emerge from the crystal on opposite sides of a cone that is centered about the direction traveled by the parent. For what is known as Type I phase matching, the daughters emerge from a specifically oriented BBO crystal with identical polarizations that are aligned perpendicular

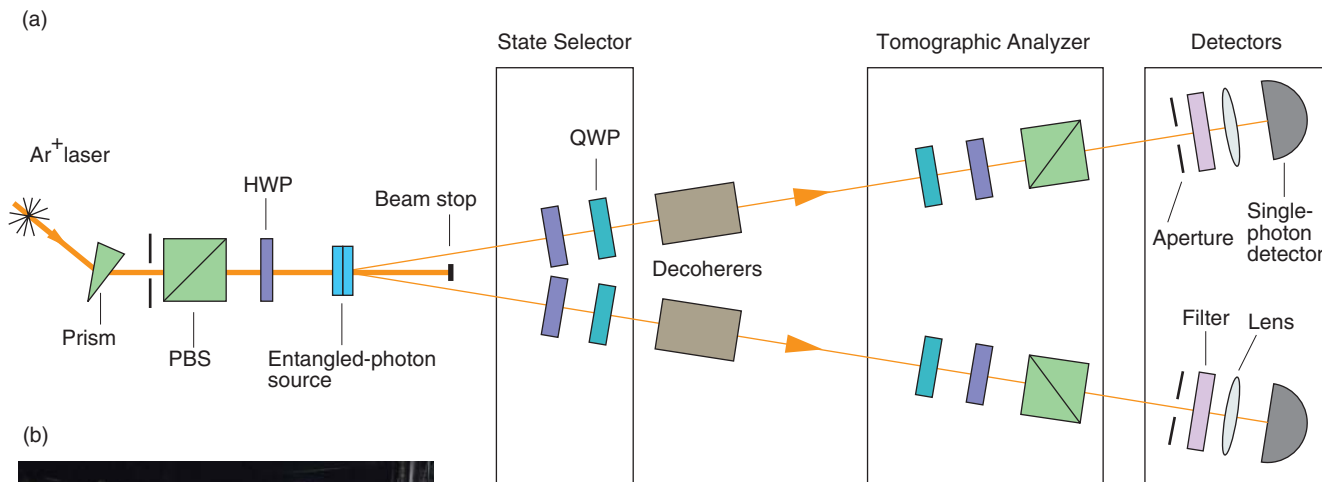


Figure 6. Creating and Measuring Two-Photon Entangled States

(a) The “parent” photons are created in an argon ion laser and are linearly polarized with a polarizing beam splitter (PBS). The half-wave plate (HWP) rotates the polarization state before the photon enters the entangled-photon source. The entangled photons produced diverge as they exit. Each photon’s polarization state can be altered at will by the subsequent HWP and quarter-wave plate (QWP). The decoherers following the state selection allow us to produce (partially) mixed photon states. The optical elements (QWP, HWP, and PBS) in the tomographic analyzer allow us to measure each photon in an arbitrary basis, for example in H/V or +45/–45. Combining the measurements on both photons allows us to determine the quantum state. (b) In the photo, Paul Kwiat is shown with the two-photon entangled source at Los Alamos.

to the parent polarization—see Figure 5(a). Because each photon is in a definite state of polarization, the two photons are not in an entangled state but are classically correlated. (The crystal acts like the source S_1 described earlier.)

To create photons in the entangled state, one can use two crystals that are aligned with their axes of symmetry oriented at 90° to each other, as shown in Figure 5(b). With crossed crystals, two competing processes are possible: The parent photon can down-convert in the first crystal to yield two vertically polarized photons ($|VV\rangle$), or it can down-convert in the second to yield two horizontally polarized photons ($|HH\rangle$). It is impossible to distinguish which of these processes has

occurred. Thus, the state of the daughter photons is a coherent quantum-mechanical superposition of the states that would arise from each crystal alone; the crossed crystals produce photons in the state $|\Psi_{\text{out}}\rangle = 1/\sqrt{2}(|HH\rangle + |VV\rangle)$, which is maximally entangled.⁴

Figure 6 shows how this basic source can be adapted to produce any pure quantum state of two photons by placing rotatable half- and quarter-

wave plates (which can be used to transform the polarization state of a single photon) before the crystal and in the paths of the two daughter photons. To create more general quantum states—mixed states—a long birefringent crystal can be used to delay one polarization component with respect to the other. If the relative delay is longer than the coherence time of the photons, the horizontal and vertical components have been effectively decohered; that is, the phase relationship between the different states is destroyed.

Researchers are still discovering how to combine sources and polarization-transforming elements to create all possible two-photon quantum states.

Characterizing Entanglement:

⁴ In an alternative approach known as “Type II phase matching,” only one crystal is needed to create the entangled state. The crystal has a different orientation, and each of the daughter photons emerges from the crystal on one of two possible exit cones. Entangled photons created by this approach were used in the first demonstration of quantum teleportation.

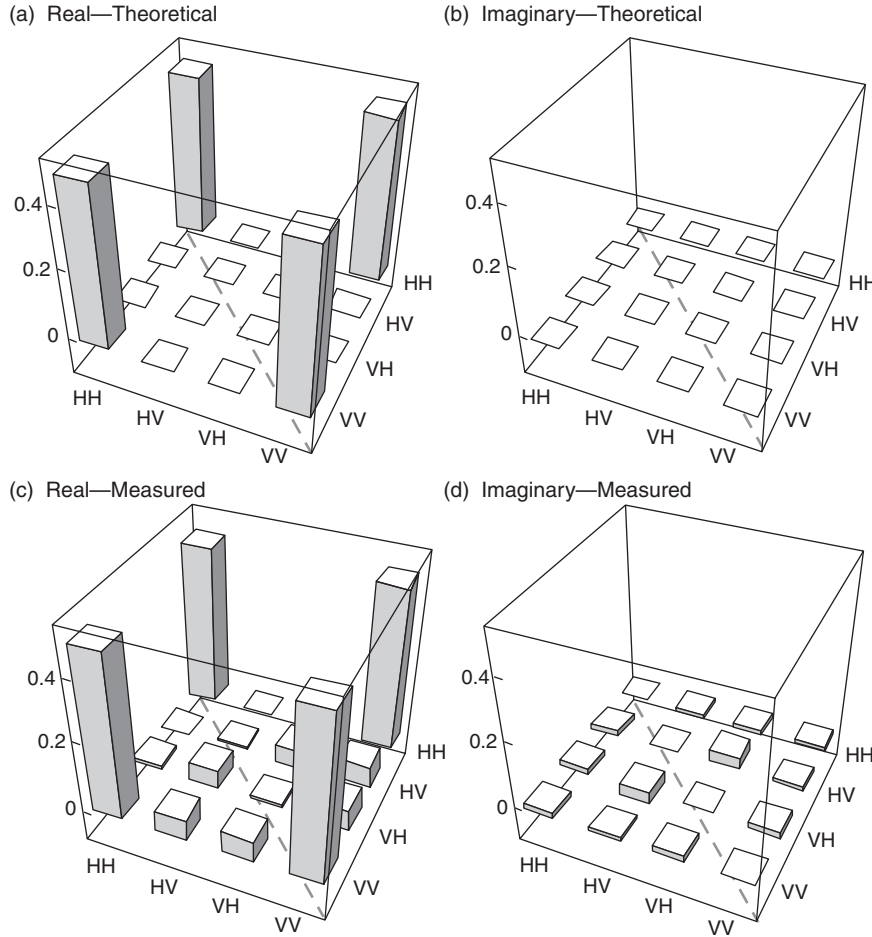


Figure 7. Density Matrices

Theoretical and experimental density matrices for the entangled state $|\Phi_+\rangle = 1/\sqrt{2} (|HH\rangle + |VV\rangle)$ are illustrated here. Both real and imaginary parts of the matrix are shown. The value of each matrix element is derived from the results of thousands of two-photon correlation experiments (simulated experiments for the theoretical matrix.) The experimental matrix indicates that our source can output a state close to a maximally entangled one. Written out “longhand,” the density matrix describing the state $|\Phi_+\rangle$ is

$$\rho = |\Phi_+\rangle\langle\Phi_+| = 1/2 (|HH\rangle\langle HH| + |VV\rangle\langle VV| + |HH\rangle\langle VV| + |VV\rangle\langle HH|) .$$

The first two terms, which lie on the diagonal of the matrix (dashed line), give the probability of the result (for example, 50% HH and 50% VV). The other two terms describe the quantum coherence between the states $|HH\rangle$ and $|VV\rangle$. For a classical mixed state (such as the source S_2 described in the text), these off-diagonal terms in the density matrix would equal zero. Notice that all coefficients in this density matrix are real, so that all terms in the imaginary part of the matrix should be zero.

The Map of Hilbert Space

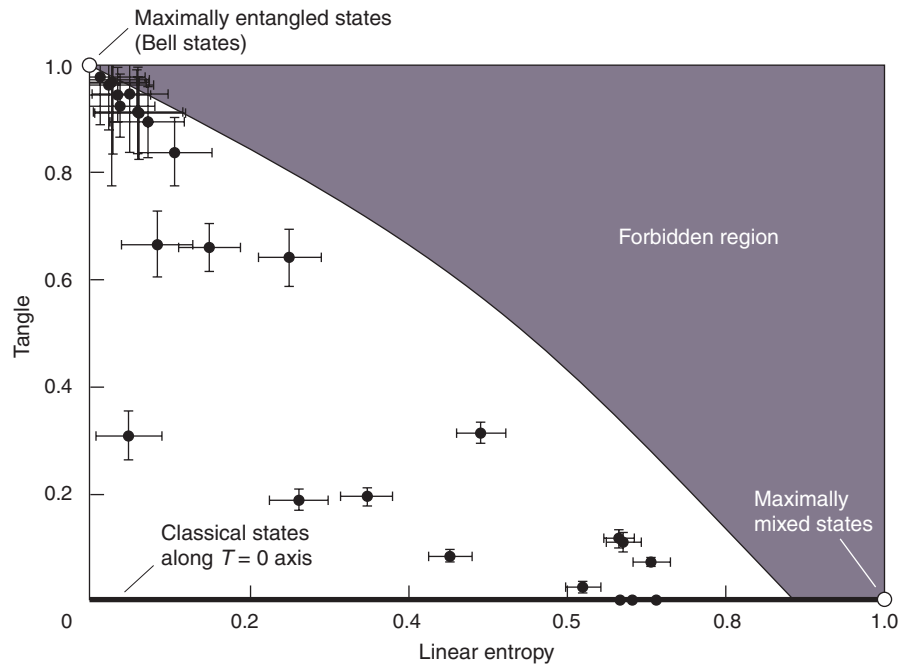
As discussed in the box on page 56, a mixed state of two photons (or in general, a mixed state of two qubits) is represented by a 4×4 density matrix, which is described by 15 independent parameters (15 real numbers). To determine the independent parameters, we make 15 coincidence measurements on the ensemble of photon pairs emitted from the source. Each measurement is similar to the one used in the simple experiment described at the start of this article. The measurement may be made with the tomographic analyzer shown in Figure 6. Using such a system, we were able to determine the density matrices of many types of states. An example is shown in Figure 7.

Whereas 15 numbers fully describe a two-photon mixed state, the density matrix for N photons needs $4^N - 1$ real numbers. Thus, the density matrix of a 4-photon state contains 255 parameters and requires 255 separate measurements just to characterize the state. Note that, if each parameter is allowed to assume one of, say, 10 possible values, those 4 photons can be in any of 10^{255} distinct quantum states! This number of states is many orders of magnitude greater than the total number of particles in our universe. The mathematical space in which the quantum states rest (the Hilbert space) is unfathomably large, and in order to have any hope of navigating it, one needs to introduce a simpler representation for quantum states.

Two characteristics of central importance for quantum information processing are the extent of entanglement and the degree of purity of an arbitrary state. A quantity called the von Neumann entropy has been introduced to characterize the degree of purity. (See the box “Characterizing Mixed States” on the next page.) However, for the analysis of two-photon states, we found it easier to use a related quantity, known as the linear entropy. When the linear entropy equals zero, the state is pure. When it reaches its maximum value of 1, the state is completely random.

Measuring the entanglement of a mixed state is more complicated and, in general, is an unsolved research problem when more than two qubits

Figure 8. The Map of Hilbert Space
The amount of entanglement (or the tangle) is plotted against the degree of purity (represented by the linear entropy) for a multitude of two-photon states created and measured at Los Alamos. Each state is represented by a black spot with error bars. The boundary line, which represents the class of states that have the maximum possible entanglement for a given value of the linear entropy, was first determined theoretically but then confirmed by a numerical simulation of two million random density matrices. Important states, such as those that are maximally entangled or completely mixed, are indicated. Efforts are under way to create states that lie along the boundary line.



are involved. Any mixed quantum state can be thought of as an incoherent combination of pure states: The system is in a number of possible pure states, each of which has some probability between 0 and 1 associated with it (rather than the complex numbers defining the probability amplitudes that specify a particular superposition of pure states). A reasonable measure of the entanglement of such a mixed state is to take the average value of the entanglement (for example, as measured by the concurrence discussed in the box on this page) for all those pure states.

One must, however, use this procedure carefully because the decomposition of the mixed state into an incoherent sum of pure states is not unique. For this “average entanglement” to make any sense as a measure of entanglement of the mixed state, one must use the decomposition for which the average is a minimum. The square of this minimized quantity is called the “tangle.” It has a value of zero for entirely unentangled, separable states and of unity for completely entangled states.

Figure 8 shows how those two

Characterizing Mixed States

It is convenient to characterize the extent of entanglement and the degree of purity of a mixed state using two derived parameters: the tangle and the linear entropy. The linear entropy, which gives a measure of the purity of the state, derives from the von Neumann entropy. The latter is given by the formula $S = -\text{Tr}\{\rho \log_2(\rho)\}$, where ρ is the density matrix. Here $\text{Tr}\{M\}$ is the trace of a matrix (that is, the sum of terms on the diagonal) and \log_2 is a logarithm base 2, which can be defined for matrices via a power series. The von Neumann entropy is zero for a pure state. When the von Neumann entropy has its maximum value (equal to the number of qubits), the state is completely random, with no information or entanglement being present. The linear entropy, defined for two qubits as $S_L = 4/3(1 - \text{Tr}\{\rho^2\})$, is similar to the von Neumann entropy, but it is easier to calculate. Specifically, it equals 0 for a pure state and has a maximum value of 1 for completely random states.

Characterizing the degree of entanglement is more difficult. Mathematically speaking, if one decomposes the density matrix into an incoherent sum of pure states, that is, $\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|$, where $0 \leq p_i \leq 1$ and $\sum_i p_i = 1$, then the average entanglement is $\bar{E} = \sum_i p_i C(\psi_i)$, where $C(\psi_i)$ is the concurrence of the pure state $|\psi_i\rangle$ (defined in the box on page 56). It is very important to find the decomposition for which \bar{E} takes its minimum possible value; otherwise, one can infer a nonzero entanglement for states such as the completely mixed state, which is certainly not entangled! Fortunately, the way to do that decomposition has been worked out for two qubits. Characterizing the degree of entanglement for three or more qubits remains an unsolved research problem.

parameters—tangle and linear entropy—can be used to create a simplified map of Hilbert space for two-photon states. The crosses (with error bars) are the states we have created and measured experimentally. Most display a high degree of entanglement. States created by other technologies can be plotted on such a diagram as well.

Conclusions

Entangled states arise naturally whenever two or more quantum systems interact. In fact, one of the prevalent theories of nature is that the universe is really one big, vastly complicated entangled state, described by the “wave function of the universe.” Despite their seeming ubiquity, however, entangled states are not generally observed in the world at large. Only relatively recently have scientists developed the means to controllably produce, manipulate, and detect this most bizarre quantum phenomenon. Initially, the fascination was limited to experimental studies of the foundations of quantum mechanics, especially the notion of nonlocal “spooklike” influences (to quote Einstein). However, even more recently, has come the realization that entanglement could lead to enhanced—sometimes vastly enhanced—capabilities in the realm of information processing.

This paper has discussed how entangled states could be a key resource in applications as diverse as cryptography, lithography, and metrology because they enable feats beyond those possible with classical physics. In addition, the quest to create a quantum computer has pushed entangled systems to the forefront of quantum research. Part of the power of a quantum computer is that it creates entangled states of N qubits so that information can be stored and

processed in the 2^N -dimensional qubit space. Quantum algorithms have been developed that would manipulate the complex entangled state and make use of the nonclassical correlations to solve problems more efficiently than could be done classically. Scientists who work on developing quantum computers are envisioning systems of thousands of entangled qubits.

We don't know whether we will be able to create or maintain such a complex entangled state. At this point, we won't even claim to know whether we will fully understand that state if it is created. More research is needed before those questions can be answered. All that we can say now is that the once-hidden domain of quantum entanglement has broken into our classical world. ■

Further Reading

- Bouwmeester, D., A. K. Ekert, and A. Zeilinger, eds. 2000. *The Physics of Quantum Information*. Berlin: Springer-Verlag.
- Haroche, S. 1998. Entanglement, Decoherence and the Quantum/Classical Boundary. *Phys. Today* **51** (7): 36.
- James, D. F. V., P. G. Kwiat, W. J. Munro, and A. G. White. 2001. Measurement of Qubits. *Phys. Rev. A* **64**: 052312.
- Mandel, L. and E. Wolf. 1995. *Optical Coherence and Quantum Optics*. Cambridge: Cambridge University Press.
- Naik, D. S., C. G. Peterson, A. G. White, A. J. Berglund, and P. G. Kwiat. Entangled State Quantum Cryptography: Eavesdropping on the Ekert Protocol. 2000. *Phys. Rev. Lett.* **84**: 4733.
- Nielson, M. A., and I. L. Chuang. 2000. *Quantum Computation and Quantum Information*. Cambridge: Cambridge University Press.
- Schrödinger, E. 1935. Discussion of Probability Relations Separated Systems. *Proc. Cambridge Philos. Soc.* **31**: 555.
- White, A. G., D. F. V. James, P. H. Eberhard, and P. G. Kwiat. 1999. Nonmaximally Entangled States: Production, Characterization, and Utilization. *Phys. Rev. Lett.* **83** (16): 3103.
- Zeilinger, A. 2000. Quantum Teleportation. *Sci. Am.* **282** (4): 50.

Daniel F. V. James was born in rainy Manchester, in northwest England, within sight of the Old Trafford (home of Manchester United, the most successful sports team in recorded human history). He was educated at New College, University of Oxford, England, and at the Institute of Optics, University of Rochester, New York, where he earned his Ph.D. in optics in 1992. He came to Los Alamos National Laboratory as a postdoctoral fellow in 1994. Three years later, he became a staff member in the Atomic and Optical Theory Group at Los Alamos. Daniel specializes in theoretical optical physics, including coherence theory, diffraction, scattering, statistical optics, and quantum technologies.



Paul G. Kwiat comes from Ohio (not a place noted for good sports teams). He was educated at the Massachusetts Institute of Technology and the University of California at Berkeley, where he received his Ph.D. in physics. In 1995, after having completed a Lise Meitner postdoctoral fellowship at the University of Innsbruck, in Austria, Paul came to Los Alamos National



Laboratory as an Oppenheimer Fellow and became a staff member in 1998. In January 2001, he became the John Bardeen Professor of Physics and Electrical Engineering at the University of Illinois at Urbana-Champaign. He specializes in experimental quantum optics, with an eye toward shedding light on quantum information protocols.

The battle between cryptographers, who encrypt messages, and cryptanalysts, who break those codes, has raged for centuries. As quantum computing promises to help cryptanalysts break many of the encryption methods used today, quantum cryptography promises to keep our secrets safe forever.





A New Face for Cryptography

Jane E. Nordholt and Richard J. Hughes

Cryptography, the mathematical science of secret communications, has had a long and distinguished history dating back to the time of the ancient Greeks. It is a subject noted for the never-ending struggle for one-upmanship between code makers and code breakers, a struggle in which the future of nations has literally been at stake. The code breakers' need to read another party's secret communications has been a tremendous force driving the development of new information-processing technologies. The code makers have responded by using those new technologies to develop more complex methods for ensuring the security of communications.

The latest round in this struggle seems set to be played out in the world's physics laboratories, with the combatants drawing upon fundamental principles of quantum physics, principles that were only of academic interest until about 15 years ago. The code breakers believe that a large-scale quantum computer—a device that uses the nonclassical aspects of quantum systems to manipulate information—could defeat the most widespread cryptosystems in use today. They are pushing the physics community to develop such a computer,

which necessarily involves controlling atoms and photons in ways that were barely dreamed of—until recently. Meanwhile, the code makers are ready for battle and are already exploiting quantum mechanics in a new code-making technology—quantum key distribution (QKD)—that could counter the quantum computing threat.

Classical Cryptography

The main goal of cryptography is to allow two parties (conventionally referred to as “Alice” and “Bob”) to communicate while simultaneously preventing a third party (“Eve”) from understanding those communications. Alice and Bob's messages should remain secret even when Eve is able to passively monitor the exchanges. (A more intrusive Eve might want to prevent Alice and Bob from communicating at all, but such a denial-of-service attack is a different type of communication problem that we will not consider here.) Cryptography provides Alice with the means to render her messages to Bob in a form that is indistinguishable from random noise but that, nevertheless, allows Bob to recover the original message.

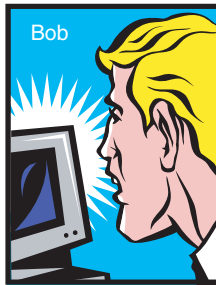
Figure 1. A Symmetric-Key Cryptography System: The One-Time Pad

(a) Alice, the sender, first generates a string of binary bits (the key) that is as long as her binary message. Then she applies the XOR operation—bit by bit—to the key and her message, and sends the encrypted string to Bob over an open communications channel. (b) Bob, the receiver, uses the same key as Alice to decrypt the message by the same XOR operation, applied bit by bit. His decrypted message is identical to the original message sent by Alice. Because the value of each key bit is random, the message cannot be recovered without the key. As long as Alice and Bob use the key only once to encrypt and decrypt one message, this one-time pad system is absolutely secure, but distributing the secret keys remains a problem. (c)–(e) This series of photographs shows an aerial view of the St. Louis International Airport before encryption, as encrypted by Alice, and as decrypted by Bob. Whereas Alice’s encrypted photo is indistinguishable from random noise, Bob is able to reproduce the original faithfully.

(a) Encryption, One-Time Pad



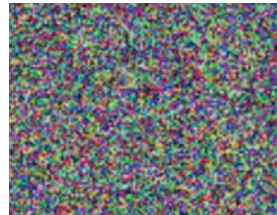
(b) Decryption



(c) Original



(d) Encrypted

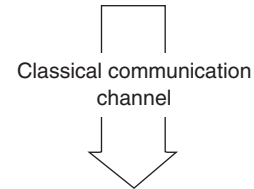


(e) Recovered Original



XOR operation, \oplus :
 $0 \oplus 0 = 0$: $0 \oplus 1 = 1$: $1 \oplus 0 = 1$: $1 \oplus 1 = 0$

Alice's message	1001	0000	0110	1001
\oplus				
Key	1000	0100	0101	0001
Encrypted message	0001	0100	0011	1000



Encrypted message	0001	0100	0011	1000
\oplus				
Key	1000	0100	0101	0001
Original message	1001	0000	0110	1001

This process of encryption (by Alice) and decryption (by Bob) can be accomplished if the two parties share a string of randomly generated binary bits known as a cryptographic key. In a system called the “one-time pad,” Alice and Bob must have identical copies of the key. (How they get the key will be discussed later). As seen in Figure 1, Alice adds the key to her message, bit by bit, using the binary exclusive OR- (XOR-, \oplus) operation, which is equivalent to addition modulo 2. Mathematically, the XOR operation is defined as

$$\begin{aligned}
 0 \oplus 0 &= 0 \quad , \\
 0 \oplus 1 &= 1 \quad , \\
 1 \oplus 0 &= 1 \quad , \text{ and} \\
 1 \oplus 1 &= 0 \quad .
 \end{aligned}
 \tag{1}$$

Alice’s encrypted communication at this point is indistinguishable from

random noise. Alice sends this message to Bob, who takes his copy of the key and subtracts it from the message, again using an XOR-operation. The original script is recovered. Provided a key is used to encipher only one message, the one-time pad encryption process is provably secure. In fact, it is the only completely secure cryptographic system.

The one-time pad is an example of a symmetric-key system (symmetric because Alice and Bob have the same key), and it requires a key that is as long as the message. In another type of symmetric key system, Alice and Bob use a short key to seed a high-quality random number generator of which they have identical copies. They then need to share fewer initial key bits in order to encrypt and decrypt large messages. In the Data Encryption Standard (DES)—a sym-

metric-key algorithm that was adopted as a United States government standard in 1977—the key length is 56 bits.

The security of all symmetric-key cryptographic systems rests entirely on the secrecy of the shared key because the structure of the cryptographic algorithm used by Alice and Bob is public knowledge. Certainly, the eavesdropper Eve understands and can implement the decryption algorithm. Should Eve obtain the key, she could immediately read Alice and Bob’s messages. Without the key, Eve must attempt a mathematical attack on the encrypted message (or parts thereof) in order to crack it. In a properly designed symmetric-key cryptosystem, no attack should be more efficient than an exhaustive search over all possible keys.

Consider, for example, the 56-bit

DES key. Because there is a choice of either 0 or 1 for every bit in a binary key, there are 2^{56} (or nearly 10^{18}) possible DES keys. A desktop computer testing a million keys a second would require more than two thousand years to search the entire key space. But the phenomenal increase in computational speed and capability has made the 56-bit key vulnerable. Today's supercomputers can search all possible keys in a matter of hours.

The simple solution is to use longer keys. Adding a bit to the key length doubles the search time, whereas doubling the key length makes the search problem exponentially harder. In the forthcoming Advanced Encryption Standard (AES), the key length will be up to 256 bits, in which case a search of the entire key space would be so computationally demanding that it would not be feasible on any computer system within the useful lifetime of the information.

The Key Distribution Problem.

A DES-type cryptographic system reduces the act of communicating a long secret (the message) to that of creating and sending a short secret (the key). But the central issue within any system is that any information about the key must remain out of the hands of unwanted parties. This latter requirement creates what is known as the key-distribution problem.

Traditionally, cryptographic keys were distributed by trusted couriers immortalized in spy movies as strangers in trench coats handcuffed to locked briefcases. But the infrastructure required to manage the key material makes this type of distribution impractical in our computer-driven, global community. Picture the logistics nightmare if a courier had to deliver a cryptographic key every time Alice wanted to use her credit card over the Internet—and imagine the added cost! In some cases, courier

key distribution is even impossible, such as when Bob is not a person but a satellite in Earth's orbit. Furthermore, the existence of the key material before delivery by courier introduces an insider threat, in that the key material could be copied and delivered surreptitiously to Eve.

About 30 years ago, researchers at Britain's Government Communications Headquarters (GCHQ), and later (independently) in the United States, found a new, more convenient way to securely distribute cryptographic keys. The system is known generically as public-key cryptography. One public-key protocol begins when Bob generates two very large prime numbers, p and q , which are multiplied to form the especially large number N . He then selects an integer g , and uses the numbers p , q , and g to generate a fourth number, d . The two numbers (N, g) constitute Bob's public key, which he makes widely available. The number d constitutes Bob's private key, which he keeps secret. (The protocol is discussed in greater detail in the box "Public-Key Cryptography: RSA" on the next page.)

When Alice wants to send an encrypted message to Bob, she grabs a copy of his public key and uses it in an algorithm that mathematically scrambles her communication. The algorithm, however, is a clever one-way operation: Bob's public key (N, g) cannot be used to unscramble Alice's encrypted message. Instead, one needs the secret number d from Bob's private key to decrypt. Given only N , it is extremely difficult to find the prime factors p and q that are needed to generate d ; hence, the system is considered secure.

Because the public-key cryptography system is asymmetric—only Bob needs to have a secret key—it has become the enabling technology for electronic commerce. Alice can grab the public key from the Bob.com

website and safely encrypt and send her credit card number. In addition, public-key encryption also provides a means for Alice to authenticate her transaction.

But public-key cryptography has its downside. Because of the computational difficulty in calculating asymmetric keys, Alice and Bob use it only to produce and distribute a symmetric key that they then use for the bulk of their discussions. More disturbing is the lack of proof that the methodology is secure. A clever person could come up with a new factoring algorithm that allows finding the secret number d , thus making public-key cryptography obsolete.

In 1994, Peter Shor of AT&T did invent such an algorithm. If implemented, that algorithm would undermine the public-key cryptography in use today. Fortunately, Shor's algorithm must be run on a quantum computer, which is currently unavailable and will probably remain so for many years.

Public-key cryptography clearly has a place where security need not be guaranteed to last for years. Because it is not provably secure, however, and because a quantum computer may render it useless in the future, a better system is needed for highly valuable data such as government or trade secrets. That better system is quantum cryptography.

Quantum Cryptography

Quantum cryptography is a type of symmetric-key distribution that allows Alice and Bob to create and share a secret key, while Eve is prevented from obtaining any more than a tiny fraction of one bit of information about the final key's binary sequence. The secret key can actually be used in any symmetric encryption method desired. Because quantum cryptography is used to send these

Public-Key Cryptography: RSA

Public-key cryptography is an asymmetric key-distribution system, wherein Bob generates two keys: a public key, which he makes available to anyone, and a private key, which he keeps secret. Alice uses the public key to encrypt her message, which she then sends to Bob, who uses his private key to decrypt that message. Perhaps the most widely used public-key cryptography algorithm is RSA, which was invented in 1978 by Ron Rivest, Adi Shamir, and Leonard Adleman and was named for its inventors. The RSA algorithm uses two keys that are constructed as follows:

- Bob generates two prime numbers, p and q , which are typically very large (several hundred bits in length).
- He calculates the product, $N = pq$, known as the modulus.
- He calculates Euler's quotient function $\Phi(N)$, which is simply the number of integers less than N that are coprime* to N . If p is a prime number, every number less than p is coprime to it, so $\Phi(p) = p - 1$. Since the modulus $N = pq$ is the product of prime numbers, $\Phi(N) = (p - 1)(q - 1)$. Let $\Phi(N)$ be designated by η .
- Bob chooses an integer g such that $g < N$, and g has no factors in common with η .
- Bob calculates $d = g^{\Phi(\eta)-1} \bmod \eta$, where $\bmod \eta$ is the modulus operation.†

Bob's public key is (N, g) . His private key is the number d .

* Two integers are coprime if they share no common divisors except 1.
 † For an introduction to modular arithmetic, see the article "From Factoring to Phase Estimation" on page 38.

When Alice wants to send a message to Bob, she first represents her message as a series of numbers. To encrypt, she grabs Bob's public key (N, g) and uses it in the following mathematical transformation:

$$c = m^g \bmod N, \tag{1}$$

where m is a number representing a piece of her message. She sends the new number c off to Bob, who uses his private key (N, d) to perform the operation

$$m = c^d \bmod N, \tag{2}$$

thereby recovering Alice's number.

Public-key cryptography is based on a theorem by Euler, which states that $x^{\Phi(y)} = 1 \bmod y$, for any integer x that is coprime to the number y . The number d was chosen such that $d = g^{\Phi(\eta)-1} \bmod \eta$, or $dg = g^{\Phi(\eta)} \bmod \eta$, which by Euler's theorem becomes $dg = 1 \bmod \eta$. Subtracting 1 will result in $dg - 1 = 0 \bmod \eta$.

The last statement indicates that the number $dg - 1$ is evenly divisible by η , so that $dg - 1 = k\eta$, where k is an integer. In decrypting the message, Bob has

$$\begin{aligned} c^d \bmod N &= (m^g)^d \bmod N, \\ &= m (m^{dg-1} \bmod N), \text{ and} \\ &= m (m^{k\eta} \bmod N). \end{aligned} \tag{3}$$

But $\eta = \Phi(N)$. By Euler's theorem, $m^{\Phi(N)} = 1 \bmod N$. Thus,

$$\begin{aligned} c^d \bmod N &= m (1)^k \bmod N, \text{ and} \\ &= m \bmod N. \end{aligned} \tag{4}$$

In other words, $c^d \bmod N = m$, so that the decryption algorithm recovers Alice's message.

key bits, it is more correctly called quantum key distribution (QKD). Adding to the security of a QKD system is the fact that any attempt to steal or copy a key can be detected, thus revealing information about the security environment.

The quantum part of quantum cryptography comes from the transmission and reception of single photons. In addition to keeping

an eavesdropper at bay (primarily because a photon cannot be split or copied reliably), quantum cryptographic systems exhibit strange quantum mechanical behaviors that are not normally observed in the classical world of everyday experience. The best example of such behavior occurs in our fiber-based quantum cryptographic system, in which we use the interference of

single photons with themselves to transmit information.

Before describing how a photon interfering with itself helps us encrypt messages, we will present an overview of the steps involved in executing a secure exchange of messages and then describe a simple protocol. Protocols are the rules used for the quantum mechanical and conventional transmissions at

the heart of QKD.

A QKD Session. To perform QKD, Alice and Bob communicate in two different ways. The first is over a quantum channel, which allows Alice to reliably send single photons to Bob. While Eve may attempt to breach the quantum channel, her tampering can be detected. The second means of communication is an ordinary, public channel assumed to be monitored by Eve. Alice and Bob use this open channel to construct their secret key, implement any of several error-correction techniques, and coordinate a “privacy amplification” scheme that effectively prevents Eve from gaining any knowledge about the final key. In all, six steps are implemented in a QKD session. These are summarized in the box to the right.

As a first step, Alice and Bob authenticate their communications; that is, they verify each other’s identity. If this step is ignored, Eve can perform a “man-in-the-middle” attack and convince Alice that she is Bob, and Bob that she is Alice, in which case no form of key distribution or encryption can prevent Eve from reading all of Alice and Bob’s communications.

After authentication, Alice and Bob begin their QKD session. First, each generates a random bit stream. Alice then uses a QKD protocol, such as BB84 (discussed in the next section), that specifies how she is to encode each bit as the quantum state of a single particle. For example, she may use the specific polarization state of a single photon to encode for either a 0 or a 1. Then, Alice would send a stream of polarized photons to Bob, who follows the protocol in determining how to measure the polarization and hence deduce a bit sequence. Because of the way the protocol works, Alice and Bob can have a public conversation and select an overlapping subset of bits without revealing to each other the value of

Six Steps to a QKD Session

Authenticate. Over an open communication line, Alice confirms she is talking to Bob, and Bob confirms he is talking to Alice.

Use a quantum protocol. The protocol dictates how Alice is to encode her random bit stream as a quantum state of a single photon. Bob measures photons according to the protocol.

Construct the sifted key. Alice and Bob use an open line to discover which photons were sent and measured in the same basis. The bit values associated with that subset of photons form the sifted key.

Construct the reconciled key. Over the open line, Alice and Bob find and remove errors from the sifted key to make the reconciled key.

Construct the secret key. Alice and Bob use privacy amplification to construct a secret key from the reconciled key. An eavesdropper has essentially no information about the bits in the secret key.

Save some bits. A few secret bits are retained to enable authenticating future QKD sessions.

those bits.

For example, if Alice’s random sequence is 0111 1010 1001 and as a result of his measurements Bob obtains the sequence 1001 1100 0100, then the protocol provides a means for Alice and Bob to know—without specifically telling each other—that the fourth, fifth, eighth, and eleventh bits form a common subsequence of 1100. This subsequence is called the “sifted” key.

In the real world, hardware is noisy, and transmission media are lossy, so the sifted key will contain some errors. Alice and Bob continue their public conversation and create a “reconciled” key, in which those errors are removed. During this process, some information about the sifted key becomes available to any potential listener (Eve). But Alice and Bob can calculate the maximum information Eve could have about their reconciled key, and using privacy amplification, reduce Eve’s information to substantially less than one bit. The result is a secret key known only to Alice and Bob. The

one remaining step before closing the session is to save a few key bits and thereby have a means to authenticate the next QKD session.

The BB84 Protocol. In 1984, Charles Bennett and Gilles Brassard published a paper describing how orthogonal and nonorthogonal quantum states could be used to construct a cryptographic key. Known today as BB84, the protocol is at the heart of our experimentally realized QKD systems. In the free-space version, Alice encodes random bit values in the polarization states of photons and then sends the single photons to Bob over the quantum channel. Bob’s measurement of the photon’s polarization and subsequent communication with Alice over a public channel allow the two parties to construct a sifted key.

A stylized version of the BB84 protocol is shown in Figure 2. (The box “Photons, Polarizers, and Projections” on page 76 also provides some background material for this section.) Alice generates a random sequence of bits and then chooses—

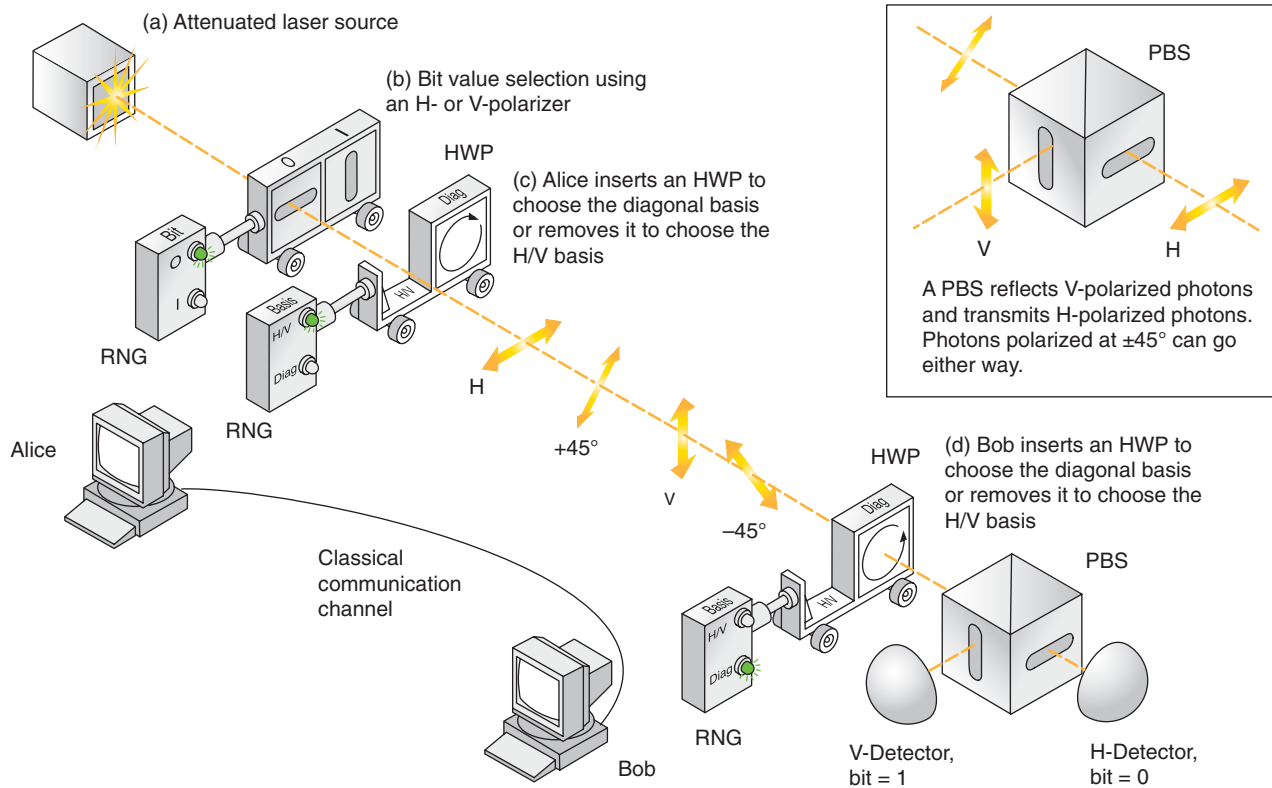


Figure 2. The BB84 Protocol

The BB84 protocol works because Alice randomly chooses to encode the photons in two, nonorthogonal bases. (a) An attenuated laser produces close to single photons. (b) Alice uses a random number generator (RNG) to select a bit value: 0s are encoded as horizontally polarized photons and 1s as vertically polarized photons (c) A second RNG selects the basis. To choose the H/V basis, Alice does nothing, (the photons are already either $|H\rangle$ or $|V\rangle$). To choose the diagonal ($-45^\circ/+45^\circ$) basis, she inserts a half-wave plate (HWP) that rotates the polarization by -45° , so that $|H\rangle$ goes to $|+45^\circ\rangle$ and $|V\rangle$ to $|-45^\circ\rangle$. (d) Bob uses an RNG to select his measurement basis, choos-

ing either to do nothing (H/V) or to rotate the photon by $+45^\circ$ ($-45^\circ/+45^\circ$). He detects photons using an H/V oriented polarizing beam splitter (PBS), which transmits horizontally polarized photons but reflects vertically polarized ones (see inset). Photons polarized at $\pm 45^\circ$ have an equal probability to go to either detector. Table I shows that, when Alice and Bob choose the same basis, they know that their bit values coincide. When they choose different bases, their bit values are randomly correlated. At the end of the session, Bob and Alice openly compare their bases for each measurement. They keep only those bits that were sent and measured in the same basis.

Table I. Details of the BB84 Protocol

Sender (Alice)			Receiver (Bob)					Joint Action
Alice's Basis	Bit	Polarization	Bob's Basis	Resulting Polarization	Probability (%)		Bit	
					H-Det.	V-Det.		
H/V	0	H	H/V	H	100	0	0	Keep bit
H/V	1	V	H/V	V	0	100	1	Keep bit
H/V	0	H	Diag.	$+45^\circ$	50	50	0 or 1	Discard bit
H/V	1	V	Diag.	-45°	50	50	0 or 1	Discard bit
Diag.	0	-45°	H/V	-45°	50	50	0 or 1	Discard bit
Diag.	1	$+45^\circ$	H/V	$+45^\circ$	50	50	0 or 1	Discard bit
Diag.	0	-45°	Diag.	H	100	0	0	Keep bit
Diag.	1	$+45^\circ$	Diag.	V	0	100	1	Keep bit

also at random—between one of two polarization bases, either the horizontal/vertical (H/V) basis, or the diagonal ($-45^\circ/+45^\circ$) basis. If she chooses the H/V basis, the bit values of 0 are encoded as horizontally polarized photons, and bit values of 1 are encoded as vertically polarized photons, that is, $0 = |H\rangle$ and $1 = |V\rangle$. Similarly, if she chooses the diagonal basis, 0 and 1 bit values are encoded as $0 = |-45\rangle$ and $1 = |+45\rangle$. She sends the stream of polarized photons off to Bob.

At his end, Bob chooses at random to measure polarizations in either the H/V or diagonal basis. As shown in Figure 2, he uses a special dual-detector system. If he chooses the H/V basis, then photons in the state $|H\rangle$ go through to his H-detector, while those in the state $|V\rangle$ are reflected to the V-detector. Photons in the $|-45\rangle$ or $|+45\rangle$ state go randomly to either detector. If Bob measures in the diagonal basis, then his setup directs $|-45\rangle$ photons to the H-detector, $|+45\rangle$ photons to the V-detector, and $|H\rangle$ or $|V\rangle$ photons to either detector with equal probability.

Table I shows how the results differ depending on which polarization states were sent and how they were detected. When Alice and Bob used the same basis, a photon hit on Bob's H-detector means that Alice had a bit value of 0; a hit on his V-detector, that she had a bit value of 1. If the bases differ, there is no such correspondence. Bob and Alice therefore use the public channel and simply compare the sequence of bases. They keep the corresponding bits when the bases agree and disregard the bits when they don't agree. In this way, they can build a sifted-key sequence over a public channel without ever revealing the value of the individual key bits.

Because Alice and Bob have a 50 percent chance of choosing the same basis, in an ideal implementation of BB84, half of the photons are used to create the sifted key. In prac-

tice, the efficiency is much less because the real world unavoidably introduces errors into the sifted-key sequence—polarizers are not perfect, photons do not always reach Bob, and detectors do not always fire when hit with a photon and sometimes fire on their own. Alice and Bob must check and correct their sequence for errors.

Error Correction. One example of a simple error-correction scheme is illustrated in Figure 3. Alice tells Bob the parity of each of her bytes, that is, whether the sum of each 8 bits of the sifted key is even or odd. Bob then checks the parity of his bytes. They keep those bytes that have the same parity and initiate a 20-questions-type deductive process to find the problem bit when the parity differs.¹ Because parity checks can only find an odd number of errors in a bit sequence, in practice, sifted bits are shuffled and then checked for errors several times. All errors must be eliminated to a high degree of certainty. If Alice and Bob's keys differ by even a single bit, the keys will be unusable.

Alice and Bob make their byte comparisons over the open channel, so Eve now has—at a minimum—information about the parity of each retained byte. To eliminate even this limited knowledge on Eve's part, Alice and Bob can agree to drop the last bit of each byte. In addition, they have to sacrifice some key bits to find the errors in their sequences. The reconciled key is therefore shorter than the sifted key. While undertaking the error correction process, however, Alice and Bob obtain an estimate of the bit error rate (BER), which is the number of errors they had in their sifted sequences. Alice and Bob use the BER and knowledge of the quantum mechanical and physical principles of the QKD technique to put a rigorous upper bound on the possible information that Eve may have about

their bit sequences.

Privacy Amplification. In this step, Alice and Bob do an XOR operation on sequences of bits from the reconciled key to produce fewer, but brand new, bits. The amount of compression required depends on their estimate of Eve's acquired knowledge.

For example, suppose Alice and Bob share a reconciled sequence consisting of six bits, a, b, c, d, e, and f, and they suspect that Eve knows three of the six bits. Alice and Bob make two new bits out of the original six by doing the following operation:

$$\begin{aligned} a \oplus b \oplus c \oplus d &= \text{Bit 1} \quad , \text{ and} \\ c \oplus d \oplus e \oplus f &= \text{Bit 2} \quad . \end{aligned} \quad (2)$$

Although Eve may have known three bits of the reconciled key sequence, she knows nothing about the new bits generated by privacy application. Alice and Bob can apply this procedure to reduce Eve's knowledge to less than one bit in a key that is several hundred bits long and thereby produce a completely secure key. In general, if the original sequence is n -bits long, privacy amplification will compress it to $R(n)$ bits, where

$$R(n) = -n \log_2[\zeta^2 + (1 - \zeta)^2] \quad (3)$$

and ζ is the BER.

Foiling Eve. We are now in a better position to discuss how the complete QKD session prevents Eve from gaining information about the secret key. First, Eve cannot get any information about the key over the open channel; although the BB84 protocol allows her to know which bits Alice and Bob had in common, she

¹ Bits that get transmitted correctly are valuable. Although Alice and Bob could drop all eight bits of a problem byte, it is usually worthwhile to winnow through the byte and retain as many bits as possible.

Photons, Polarizers, and Projection

Our realization of the BB84 protocol uses the polarization state of individual photons to encode bit values. But the key feature that prevents an eavesdropper from detecting the polarizations without being noticed is the use of two nonorthogonal linear polarizations to represent 0 and 1. Rather than preparing a random sequence of horizontally or vertically polarized photons in the quantum states $|H\rangle$ or $|V\rangle$, respectively, Alice (the sender) polarizes photons in the quantum states $|H\rangle$ or $|−45\rangle$ when she wants to send a 0 to Bob (the receiver) and $|V\rangle$ or $|+45\rangle$ when she wants to send a 1.

We can do a simple experiment to demonstrate the quantum mechanical properties of nonorthogonal photons. We need just 3 sheets of linearly polarizing filters, which are readily available from scientific education kits or suppliers. The filter is made from a material that has an intrinsic transmission axis for photons (the polarization axis). As shown in Figure A, if randomly polarized light (for example, sunlight), made up of a large number of photons goes through a linear polarizer with its axis aligned, say, horizontally, the photons that emerge are polarized in the state $|H\rangle$.

We perform our experiment by orienting the first polarizer filter horizontally and holding it up to sunlight. The light intensity decreases by about 50 percent, which indicates that about half the photons get through. We then place a second polarizer behind the first and rotate it until no light passes. At that point, the polarization axes of the two filters are orthogonal to each other, that is, the polarization axis of the second polarizer is in the vertical direction. If we place the third filter between the first two with its polarization axis at $−45^\circ$ to the others, we naively expect no change in the light transmission, but suddenly one eighth of the sunlight gets through the stack, even though the axes of the outer two polarizers are still perpendicular.

These spooky results are a direct consequence of the quantum properties of single photons. A linearly polarized photon is described by a quantum mechanical wave function. Mathematically, it is represented by a “ket” $|\psi\rangle$, which is analogous to an ordinary unit vector in 2-dimensions. Just as a plane vector can be written in terms of two orthogonal plane vectors, we can express $|\psi\rangle$ as a superposition of two orthogonal kets, $|\phi\rangle$ and $|\phi+90\rangle$, in a two-dimensional Hilbert space, with real (as opposed to complex) coefficients. The ket $|\phi\rangle$ represents a photon linearly polarized at the angle ϕ to the horizontal, while $|\phi+90\rangle$ represents a photon polarized at the angle $(\phi + 90^\circ)$. The orthogonal kets are a basis for the Hilbert space. We have

$$|\psi\rangle = \cos\theta |\phi\rangle + \sin\theta |\phi+90\rangle, \tag{1}$$

where θ is the angle between $|\phi\rangle$ and $|\psi\rangle$. The coefficients in front of the kets $-\cos\theta$ and $\sin\theta$ are probability amplitudes. Nature has dictated that the outcome of a measurement of the photon’s polarization state (for example, by transmission through a polarizing filter) is indeterminate—it depends on the basis (the orientation of the polarization axis) used to make the measurement. The probability p that a measurement of $|\psi\rangle$ yields the result $|\phi\rangle$ is given by the expression

$$p = \cos^2\theta, \tag{2}$$

that is, p is the square of the probability amplitude in front of the ket $|\phi\rangle$.

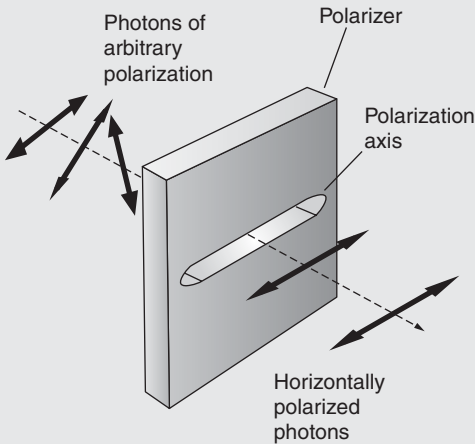


Figure A. Polarizing Filter
The filter projects photons into polarization states parallel to its polarization axis.

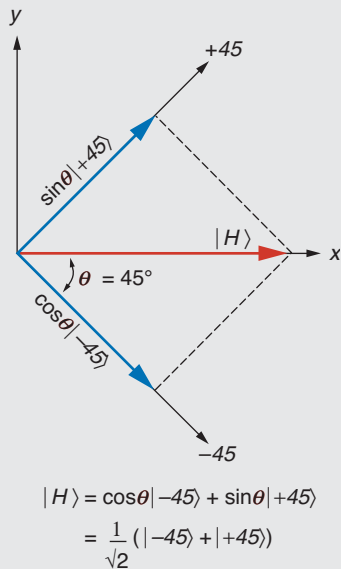


Figure B. Decomposition into Diagonal Basis
A horizontally polarized photon is expressed in terms of the $+45/-45$ basis.

We are now in a position to understand the simple experiment discussed earlier. The polarization axis of the first polarizing filter is set to be horizontal. Equation (1) tells us we can express an incoming photon as a superposition of a ket that is aligned parallel the polarization axis, that is, $|\phi\rangle = |0\rangle \equiv |H\rangle$, and a ket that is orthogonal to the axis, that is, $|\phi+90\rangle = |90\rangle \equiv |V\rangle$. We have

$$|\psi\rangle = \cos\theta |H\rangle + \sin\theta |V\rangle, \quad (3)$$

where the angle θ is now seen to describe the angle between the incoming photon's polarization and the filter's polarization axis. According to Equation (2), the probability that a linearly polarized photon passes through the horizontal polarizer is $p = \cos^2\theta$, that is, the square of the probability amplitude for the state $|H\rangle$. Because photons of all polarizations impinge on the first filter, the amount of light that gets through found by taking the average of p over all angles, that is, $\langle \cos^2\theta \rangle = 1/2$. Half the light makes it through the first filter.

Every photon that makes it through has been projected into the state $|H\rangle$. These photons then interact with the second filter in the stack with polarization axis aligned at $\phi = -45^\circ$. We express the horizontal photon in the diagonal ($-45^\circ/+45^\circ$) basis as (see Figure B):

$$|H\rangle = \cos(45)|-45\rangle + \sin(45)|+45\rangle = 1/\sqrt{2} (|-45\rangle + |+45\rangle). \quad (4)$$

The probability that a photon passes through the second filter is $\cos^2(45) = 1/2$, so 1/4 of the sunlight makes it through the two filters. The photons that emerge are polarized at -45° . The third filter is aligned vertically ($\phi = 90^\circ$), so we rewrite the ket $|-45\rangle$ in the horizontal/vertical (H/V) basis:

$$|-45\rangle = \cos(-135)|V\rangle + \sin(-135)|-H\rangle = 1/\sqrt{2} (-|V\rangle + |-H\rangle). \quad (5)$$

The probability that a photon passes through the vertical filter is $\cos^2(-135) = 1/2$. Again, half the photons make it through the last filter, so in total one eighth of the sunlight makes it through the stack.

This demonstration of nonorthogonal photon polarizations and polarizers reveals another important property of photons: All information about the initial polarization state is lost as a result of the photon-polarizer interaction. For cryptography, that has an unfortunate implication for someone (Eve) who is trying intercept the encrypted bit stream. Eve can intercept the photons going to Bob, but unless she measures the polarization of those photons in the correct basis, she cannot correlate the results of her measurements with a bit value. With her polarizer set to -45° , she has a probability to detect photons in the state $|-45\rangle$, $|H\rangle$, or $|V\rangle$, corresponding to bit values of 0, 0, and 1. Her measurement does not reveal Alice's bit value, nor does it reveal the original polarization state of the photon. A certain fraction of the photons she sends to Bob (which she must do to cover her tracks) will be in error. Thus, by choosing to send a random sequence of nonorthogonally polarized photons, Alice and Bob assure that Eve cannot attempt to measure the sequence without introducing detectable errors in their QKD protocol.

knows nothing about the values of those bits. If Eve is to get bit information, she is forced to breach the quantum channel by intercepting the photons and measuring their polarizations. She must then send new photons on to Bob in order to cover her tracks.

But Eve must know the exact state of a photon if she is to send a new one correctly. She cannot, however, make a deterministic measurement of the photon's polarization state because Alice sends photons in two nonorthogonal bases. For example, suppose Eve has a detection apparatus identical to Bob's and she detects a photon in her first detector (bit value of 0) when she measures in the diagonal basis. Did Alice send a photon in the $|H\rangle$, $|V\rangle$, or $|+45\rangle$ state? Eve has no idea because, given her measurement basis, she can detect each of those states. A hit on Eve's detector does not reveal whether Alice sent a 0 or a 1; that information "materializes" only after Alice and Bob compare bases. In fact, Eve can choose any type of detection system or measurement strategy and still be uncertain about the original state of Alice's photon.

One might ask whether Eve can make copies of Alice's photon before making a measurement. Then she could send the original off to Bob, save her string of photons (somehow), and make deterministic polarization measurements after listening to Alice and Bob compare bases. But quantum mechanics prevents Eve from accurately copying an unknown photon. (See the box "The No-Cloning Theorem" on page 79.) She would have to make a deterministic measurement, but that action would inevitably reveal her presence to Alice and Bob.

If she were to guess the polarization state, Eve would have, at best, a 50 percent chance of forwarding the correct one to Bob. But in making her



Figure 3. A Simple Error-Correction Scheme

Error correction removes single-bit errors from the sifted key. A simple scheme involves checking the parity of each byte (8-bit) sequence. The parity of a byte is 0 if the number of 1s in the byte is even or 1 if the number of 1s in the byte is odd. In this case, Alice and Bob start a public conversation to compare the parity of each of their three bytes. Because there is a mismatch, caused by the seventh bit (indicated in red) in the third byte, they try to locate the problem. They must eliminate all errors, or else their keys are unusable. Because the conversation takes place over an open communication line, Eve initially gains information about the parity of the sifted key. That information, however, can be eliminated if Alice and Bob drop some bits from their sequence. Relying on her old information, Eve will not understand anything about the new bit sequence.

guess, she will necessarily introduce errors into Alice and Bob's sifted-key sequence and, hence, increase the BER. When Alice and Bob check their sifted-key sequences for mismatches, they conservatively assume that Eve caused all the errors. They make corrections to those sequences, compute the maximum information Eve could have about the reconciled key, and then use privacy amplification to compress out Eve's possible knowledge about their

shared secret strings to substantially less than one bit. The secret key is truly secret.

Experiments

To date, the three major experiments performed at Los Alamos National Laboratory are free-space, fiber, and entangled-state QKD systems. All of the systems were constructed from readily available pieces

of equipment, and we were able to show that a complete QKD session could be communicated over long distances and still produce a useful secret-bit yield. All three systems use the BB84 protocol.

Here, we describe the free-space and fiber-based experiments. Entangled-state QKD is described in the article "Quantum State Entanglement" on page 52.

Free-Space QKD. In free-space

The No-Cloning Theorem

In 1982, Bill Wootters and Wojciech Zurek applied the linear properties of quantum mechanics to prove that an arbitrary quantum state cannot be cloned. Although their argument is entirely general, we will illustrate the theorem with polarized photons. Suppose we have a perfect cloning device in the initial state $|A_0\rangle$ and an incoming photon in an arbitrary polarization state $|s\rangle$. The device duplicates the photon as follows:

$$|A_0\rangle|s\rangle \rightarrow |A_s\rangle|ss\rangle, \quad (1)$$

where $|A_s\rangle$ is the device final state, which may or may not depend on the polarization of the original photon, and $|ss\rangle$ refers to the state of the electromagnetic field in which there are two photons, each with polarization $|s\rangle$. Suppose that the device can duplicate both the vertical $|V\rangle$ and the horizontal $|H\rangle$ polarization, that is,

$$|A_0\rangle|V\rangle \rightarrow |A_V\rangle|VV\rangle, \quad \text{and} \quad (2)$$

$$|A_0\rangle|H\rangle \rightarrow |A_H\rangle|HH\rangle. \quad (3)$$

According to quantum mechanics, this transformation should be representable by a linear operator, which means the operator acts independently on each orthogonal state in the Hilbert space. Therefore, if the incoming photon has some arbitrary polarization given by the linear superposition $|s\rangle = \alpha|V\rangle + \beta|H\rangle$, the result of its interaction with the apparatus will be a superposition of Equations (2) and (3):

$$\begin{aligned} |A_0\rangle|s\rangle &= |A_0\rangle (\alpha|V\rangle + \beta|H\rangle) \\ &= \alpha|A_V\rangle|VV\rangle + \beta|A_H\rangle|HH\rangle. \end{aligned} \quad (4)$$

If the apparatus states $|A_V\rangle$ and $|A_H\rangle$ are not identical, the two photons emerging from the apparatus are in a mixed state of polarization; if they are identical, the emerging two photons are in a pure entangled state, $\alpha|VV\rangle + \beta|HH\rangle$. In neither case does the apparatus produce a final state $|ss\rangle$ consisting of two completely independent photons, each in the polarization state $\alpha|V\rangle + \beta|H\rangle$:

$$\begin{aligned} |ss\rangle &= (\alpha|V\rangle + \beta|H\rangle) (\alpha|V\rangle + \beta|H\rangle) \\ &= \alpha^2|VV\rangle + \alpha\beta|VH\rangle + \beta\alpha|HV\rangle + \beta^2|HH\rangle. \end{aligned} \quad (5)$$

Linearity, therefore, rules out the existence of a device that could faithfully clone a photon in an arbitrary polarization state.

QKD, photons are transmitted through open air. The protocol uses polarization states, as previously described, because the atmosphere preserves polarization over a wide range of photon wavelengths (including the full range of visible and infrared light).

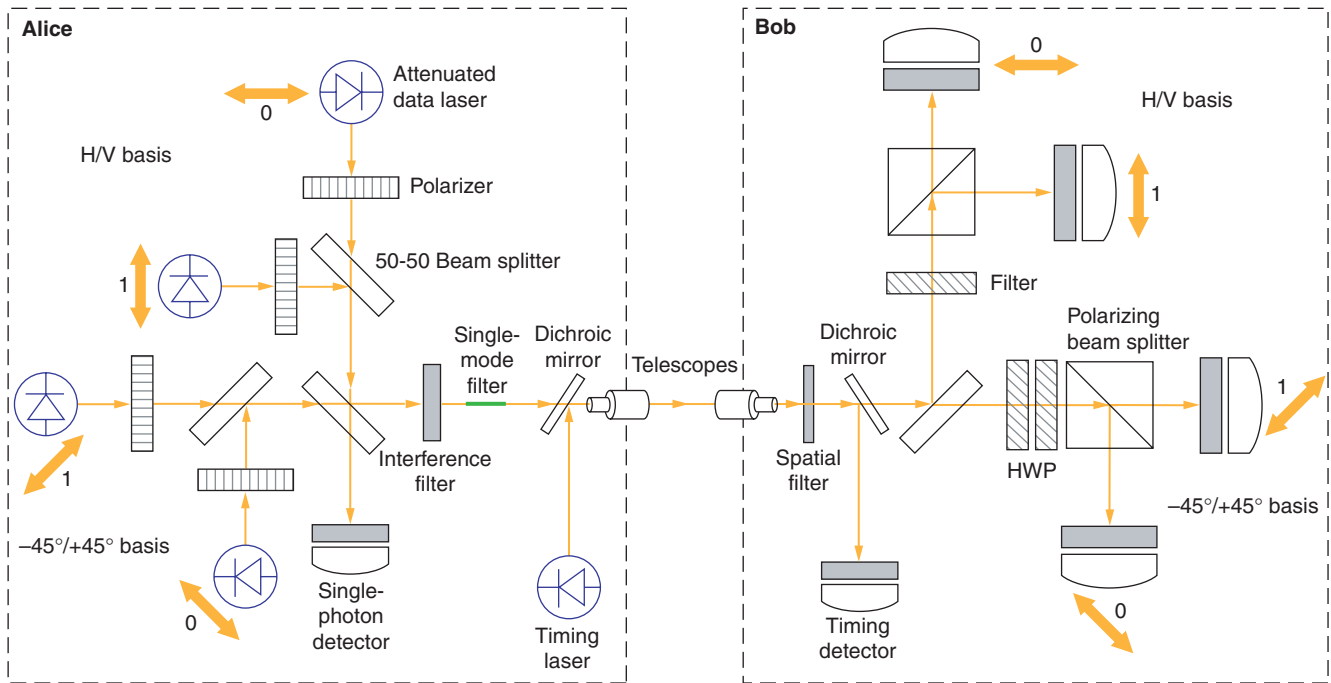
The major difficulty is detecting the single QKD photons from within the enormous background of daytime photons, namely, $\geq 10^{10}$ background photons per centimeter squared, per second, per angstrom, per steradian ($\gamma/\text{cm}^2/\text{s}/\text{\AA}/\text{sr}$). This problem exists

even at night because the background from, say, moonlight or the light of urban areas is still much larger than the QKD signal. A second difficulty is dealing with losses due to atmospheric distortions. We are able to overcome both of these problems and can distinguish the QKD photons from background photons by using interference filters that transmit only photons of a specific wavelength, by carefully limiting the field of view, and by using a clever trick. The free-space QKD system is shown in Figure 4.

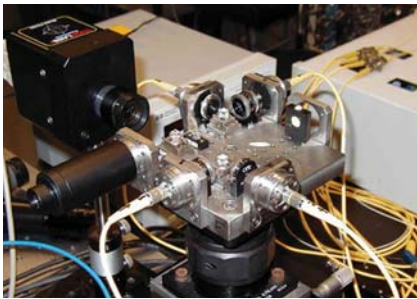
Alice and Bob have identical copies of the interference filters, which allow Alice to send photons at a selected wavelength and Bob to receive photons only at that wavelength. The preferred wavelength is about 772 nanometers, which is in the infrared and just outside the normal range of vision. The atmosphere is highly transmitting for light of this color, and single-photon detectors with good quantum efficiency at this wavelength are readily available. Furthermore, polarization selection and control components and diode lasers that produce the desired wavelength are all easily obtained.

A receiver telescope with a narrow field of view helps limit unwanted photons. Behind the telescope is a spatial filter that passes photons coming from a precise location (Alice's) while excluding all the others. The telescope must be employed with care, however. As anyone who has ever looked at the twinkling stars knows, the atmosphere can make a source of light appear to move. The magnitude of the movement varies considerably with the time of day, the weather, and the local terrain. If not accounted for, the atmosphere could cause Alice to shift rapidly in and out of Bob's field of view. Over short distances, these atmospheric distortions are not a serious prob-

(a) Conceptual Diagram



(b) Alice's Optics Table



(c) Alice's Electronics



(d) Bob's System



Figure 4. Free-Space QKD

(a) In the BB84 protocol, Alice (the sender) encodes bits in the polarization states of single photons either as $0 = |H\rangle$ and $1 = |V\rangle$ or as $0 = |-45^\circ\rangle$ and $1 = |+45^\circ\rangle$. The data stream begins with a bright output pulse from the timing laser, which sets the timing of the pulse. A few nanoseconds later, one of the four data lasers ($\lambda = 772 \text{ nm}$) fires. Each data laser has its own attenuator, focusing optics, and polarizer. Each laser outputs a uniform pulse of the desired brightness in one of the four polarization states. The output of all four data lasers is combined by a series of beam splitters, which have been carefully arranged so that the distances between the lasers and output optics are the same (therefore eliminating any timing differences between the pulses). The final beam

splitter either directs the photons to a detector that monitors the average number of photons per laser pulse or sends the polarized photons through a narrow-pass interference filter (to remove any frequency differences) and a single mode fiber (to eliminate any spatial mode differences). The photons that pass through Alice's telescope are identical in every respect except for polarization. Bob (the receiver) uses spatial filtering, time-domain filtering, and wavelength selection to pick out Alice's photons from background. His telescope, with a field of view that is nominally 45 arc seconds (or 220 microradians), acts as a spatial filter that allows only photons from Alice's location to pass. The photons then pass through an interference filter (wavelength selection)

that is matched to the one in Alice's transmitter. Photons are sent to a 50-50 beam splitter, which acts as a basis selector by randomly directing a photon to one of the two measurement stations. Each station consists of a polarizing beam splitter and two single-photon detectors. A half-wave plate (HWP) rotates the photon's polarization before the $-45^\circ/+45^\circ$ station. A detector must fire within a set period following detection of the bright timing pulse (time-domain filtering). (b) Alice's compact optics table and (c) electronics are shown here. (d) Bob's telescope peers out from the door of the mobile trailer containing all his electronics and optical systems. Bob (and Alice) can be easily transported to different sites. Moreover, one person can operate the system.

lem. Over long distances, Alice corrects for atmospheric variations by observing Bob's beacon laser and is thus able to rapidly vary the point to which she sends the photons.

Finally, the clever trick is to send a bright laser pulse from Alice to Bob just before a single photon is sent so there is a known delay between the photon and the bright pulse. Bob accepts only photons that enter the system approximately 1 nanosecond after the bright pulse. This time-domain filtering greatly limits the possibility of a background photon being detected instead of a QKD photon. This system of multiple filtering techniques works so well that single QKD photons can be distinguished from background even in daylight.

One issue complicating the free-space system (as well as the other systems described below) is that the photon sources are actually attenuated laser diodes that produce weak laser pulses instead of true single photons. (Single-photon sources are currently too large and exotic for systems intended for use in the field.) The number of photons in a weak laser pulse is governed by Poisson statistics, and the number of photons in each pulse varies. The probability $P(n)$ that a pulse will contain n photons is,

$$P(n) = \frac{e^{-\mu} \mu^n}{n!}, \quad (4)$$

where μ is the average number of photons per pulse. If $\mu = 1$, there is roughly a 37 percent chance that a pulse will contain no photons, 37 percent that it will contain one photon, and 26 percent that the pulse will contain more than one photon.

By adjusting the attenuation, Alice can choose a specific value of μ . If she chooses a relatively high μ , say, above 1 photon per pulse, each time more than one photon is sent, it

must be assumed that a clever eavesdropper would be able to detect and measure the extra photons. A great deal of privacy amplification—concomitant with a large consumption of reconciled bits—is needed to keep the system secure, so overall, the secret bit yield decreases. If μ is too small, say, 0.05, then most of the time Alice is sending nothing over the quantum channel and experimental errors (such as background light getting into the receiver, dark counts in detectors, or even the actions of an eavesdropper) may dominate. Again, the secret-bit yield decreases. The choice of μ is therefore an important free parameter at Alice's disposal.

Our experiments have shown that the secret-bit yield depends strongly on atmospheric conditions. Turbulence along the optical path between Alice and Bob, for example, affects the transmission efficiency. To help show trends in the data, we construct a pseudo signal-to-noise ratio, η/C , where η is the transmission efficiency (obtained by dividing the number of sifted bits by μ) and C is the number of background photons detected by Bob.

Figure 5 shows data from a free-space QKD experiment that ran successfully at a 10-kilometer separation in daylight. The open communication channel was a wireless Ethernet. During the numerous experimental runs, Alice would send 10^6 laser pulses over a 1-second period. The value of μ was typically set between 0.1 and 0.8.

The experimental run labeled "Sample" in Figure 5 is a typical example. Approximately 22 percent of the pulses had a single photon ($\mu = 0.29$). After comparing Alice and Bob's bases, we constructed a sifted key of 651 bits. Following error correction, calculation of the BER, and privacy amplification, we obtained a secret key consisting of

264 bits, which is sufficient for the new AES. Note that the secret-bit yield can be substantially higher at night (high η/C), because the background is reduced.

Our free-space system is a preliminary prototype for a system that could be flown on a spacecraft. Because the atmosphere has an effective thickness of only a few kilometers if one were to look straight up, our results are a good indicator of the feasibility of ground-to-satellite free-space QKD.

Fiber-Based QKD. The polarization state of a photon is not preserved in conventional optical fibers. That is why another physical property that could express the desired quantum mechanical properties for QKD had to be found in order to implement a fiber-based system.

The solution was to have a photon interfere with itself after it travels down two paths of a twin Mach-Zehnder interferometer setup.

The concepts underlying the fiber-based QKD scheme are illustrated in Figure 6. Briefly, quantum mechanics tells us that a single photon entering a Mach-Zehnder interferometer behaves as if it has taken both paths through the instrument. The entrance beam splitter places the photon in a quantum mechanical superposition, with a component that describes a photon traversing the upper path and a component that describes the photon traversing the lower path. The two components have a definite phase relationship and can interfere with each other.

As seen in the figure, Alice can introduce a phase shift ϕ_A to the photon on one arm of the interferometer, while Bob can introduce a phase shift ϕ_B on the other. Depending on the phases set by both Alice and Bob, the interference

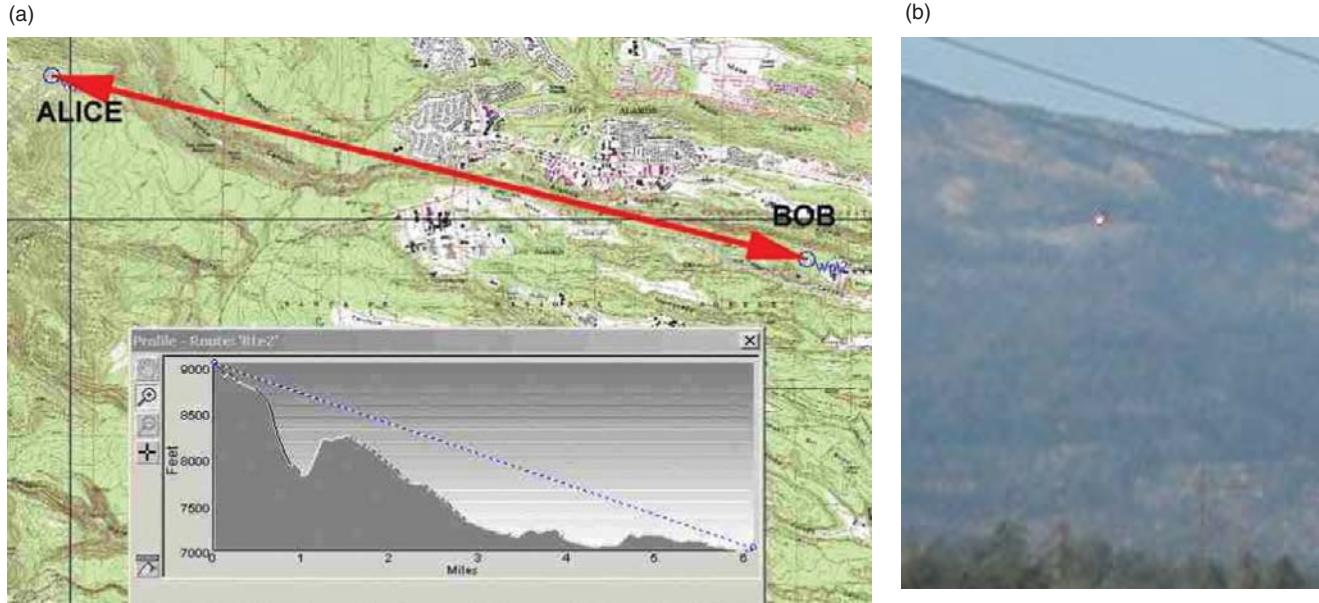
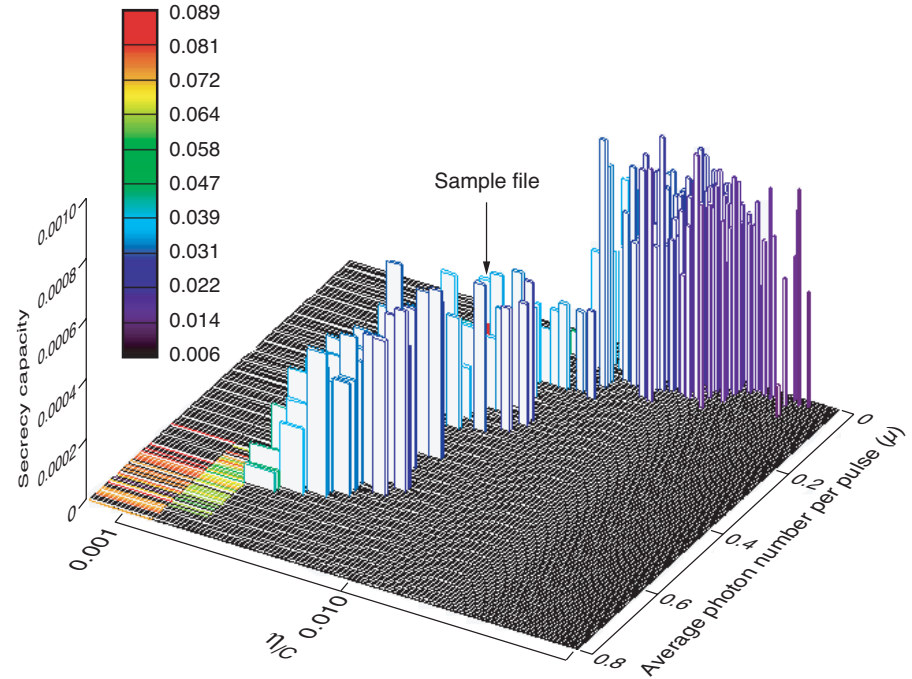


Figure 5. Data from a 10-km Free-Space QKD Experiment

(a) Alice was located halfway up Pajarito Mountain, in New Mexico, while Bob was 10 km away, at a Los Alamos lab site. (b) The bright red dot near the center of the picture is a spotting laser sent through Alice’s telescope. It was used to optically align the transmitter and receiver for the quantum channel. (c) Data from the experiment show the dependence of the secret-bit yield (normalized to the number of sifted bits) on the average number of photons per pulse μ and on the pseudo signal-to-noise ratio η/C (discussed in the text). Each vertical column corresponds to an experimental run in which Alice sent 10^6 polarized photons in 1 s. The flat, black regions of the graph are areas for which no data are available. With favorable atmospheric conditions or low background (high η/C), we can run at lower μ values and still obtain a high bit yield. Poorer conditions (low η/C) require higher μ values and result in a lesser yield.

(c) Sifted-Bit Error Rate (r)



at the exit beam splitter is such that the photon has a definite probability to hit either of two detectors. The probability P_U that the photon hits the upper detector is given by

$$P_U = \sin^2\left(\frac{\phi_A - \phi_B}{2}\right), \quad (5)$$

whereas the probability P_L that the photon hits the lower detector is given by

$$P_L = \cos^2\left(\frac{\phi_A - \phi_B}{2}\right) \quad (6)$$

We make use of these relations to implement the BB84 protocol. Alice

chooses at random between two bases, X and Y. If she chooses the X-basis, then for a bit value of 0 or 1, she sets $\phi_A = 0^\circ$ or 180° , respectively. If Alice chooses the Y-basis, then she chooses $\phi_A = 90^\circ$ or 270° for bit values of 0 or 1, respectively. At his end, Bob sets his phase angle ϕ_B to 0° if he is in the X-basis and to 90° if he is in the Y-basis.

Table II summarizes Alice and

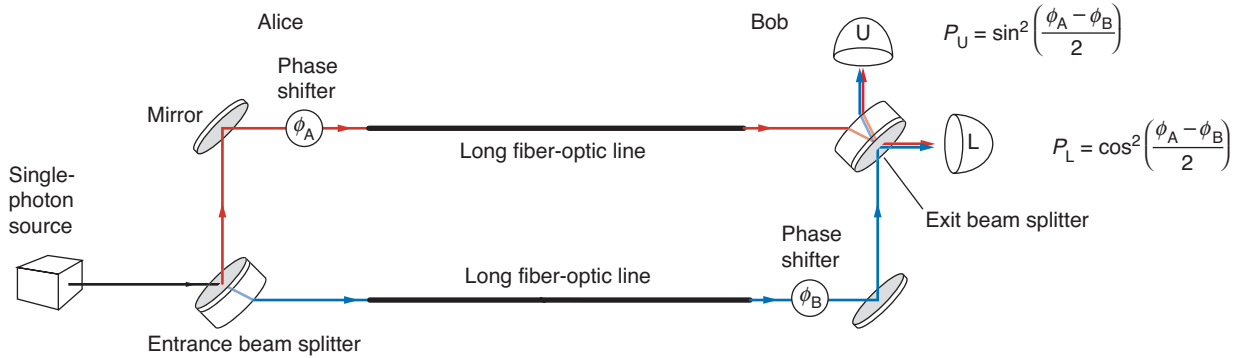


Figure 6. Mach-Zehnder Interferometer and Fiber-Based QKD Concept

In a Mach-Zehnder interferometer, a photon is placed in a superposition of two states by the entrance beam splitter. It travels down both arms simultaneously, and interferes with itself at the exit beam splitter. In the conceptual fiber-based QKD system illustrated here, a phase shifter is placed in each arm of the interferometer. Alice randomly chooses a bit value and a basis and sets the angle of her phase shifter according to her choices (see Table II below). Bob sets the

angle of his phase shifter according to his basis choice. The table shows the probability that Bob detects a photon in a given detector. When Alice and Bob use the same basis for sending and measuring, a hit in Bob's lower detector means that Alice sent a bit value of 0, whereas a hit on the upper detector means she sent a 1. Because there is no such correlation when Alice and Bob use different bases, those bit values are discarded.

Table II. Fiber-Based QKD

Sender (Alice)			Receiver (Bob)				Action	
Basis	Bit	Phase ϕ_A (°)	Basis	Phase ϕ_B (°)	Probability (%)			Bit
					P_L	P_U		
X	0	0	X	0	100	0	0	Keep bit
X	1	180	X	0	0	100	1	Keep bit
X	0	0	Y	90	50	50	0 or 1	Discard bit
X	1	180	Y	90	50	50	0 or 1	Discard bit
Y	0	90	X	0	50	50	0 or 1	Discard bit
Y	1	270	X	0	50	50	0 or 1	Discard bit
Y	0	90	Y	90	100	0	0	Keep bit
Y	1	270	Y	90	0	100	1	Keep bit

Bob's choices and shows the value of the probabilities P_U and P_L , given the various combinations of ϕ_A and ϕ_B . Because we are implementing BB84, Table II is essentially the same as Table I. When Alice and Bob choose the same basis, a photon representing Alice's 1 always goes to the upper detector, and a photon representing her 0 always goes to the lower. If Alice and Bob use different bases, the photon has equal probability to emerge from either port, and Bob has no information about what bit value

Alice has sent. At the end of the session, Bob calls Alice on the open communications line, and the two compare which bases they used for each photon. They keep the bit values when the bases agree and discard the other bits.

In the scheme discussed above, a single Mach-Zehnder interferometer stretches between Alice and Bob. In practice, that is a bad idea. The photon needs to maintain phase coherence as it propagates down the two optical fibers that make up the long arms of

the interferometer. Photons often experience random phase shifts as they go through long fiber-optic cables, and because the shifts in one arm are independent of those in the other, the interference condition at the exit beam splitter changes in a random fashion. Furthermore, having two dedicated fibers would be expensive to operate in the real world.

A better idea is for Alice and Bob each to have a Mach-Zehnder interferometer, with the two connected by a single long fiber—see Figure 7.

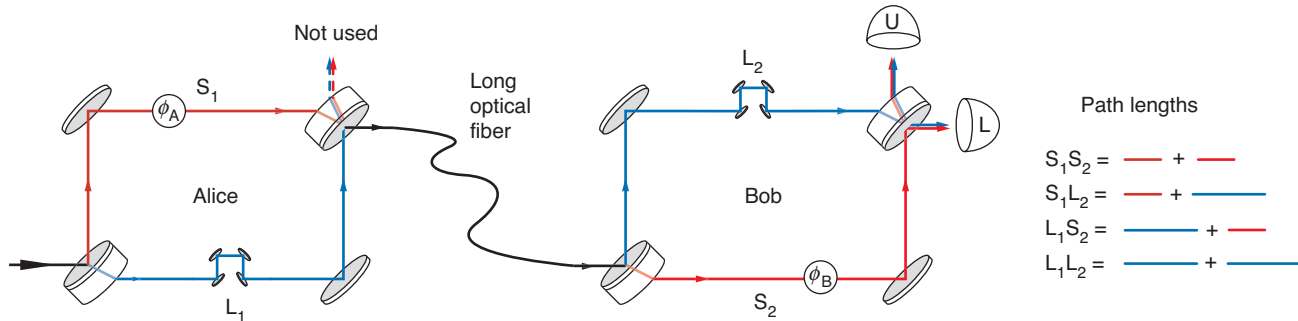


Figure 7. Implementation of Fiber-Based QKD

Our fiber-based QKD system uses two modified interferometers connected by a single, long optical fiber. Each interferometer has a long (L) arm and a short (S) arm. In going from Alice’s entrance beam splitter to Bob’s exit beam splitter, the photon can take paths S_1S_2 , L_1L_2 , S_1L_2 , and L_1S_2 . The latter two paths have the same length, and the photon traveling them can maintain phase coherence and interfere with itself. The protocol then works as described in Figure 6.

Each interferometer is modified to have a long arm and a short arm, and the path length differences between the two arms are greater than the coherence length of the photon. There is no interference as the photon leaves Alice’s instrument. But of the four possible paths through the entire system (refer to the figure), the two designated as S_1L_2 and L_1S_2 are of equal length (to within the phase coherence length of the photon). A photon that travels down those two paths interferes with itself at Bob’s exit beam splitter. The system therefore behaves as if it were a single instrument. Alice and Bob are still free to vary the phase on one arm of their interferometers, as needed, to carry out the protocol.

Our system transmits bits through 48 kilometers of fiber. As in the free-space experiments, Alice first sends a bright pulse to trigger the detectors and to limit background interference. Single photons are sent at 1310 nanometers, and the bright timing pulse is at 1550 nanometers. The secret-bit yield is lower than that obtained in the free-space experiment.

Summary

Quantum cryptography can enable secure transmission of sensitive, pro-

prietary, or national security information across a metropolitan area or corporate campus and provide the long-term security guarantees such data require. It is the only technology that will be secure no matter what technology an adversary develops in the future. Furthermore, it raises the stakes for eavesdroppers because they must perform risky, active attacks against a system. Currently, a public-key encrypted system can be attacked through passive, standoff monitoring.

Because of the inherent advantages of quantum cryptography, we can envision a future in which a QKD system provides secure communications in metropolitan areas between banks, between off-site stock-trading centers and central stock exchanges, between corporate offices, and between offices and broadband data networks. Money transfers between banks now amount to over \$2 trillion per day worldwide and well justify the expense of implementing QKD systems. Optical wireless “last-mile” communications systems could even provide broadband access to most homes.

By combining theoretical analyses with innovative experimental advances, the Los Alamos quantum cryptography team has already demonstrated the practicality of free-

space quantum cryptography in a series of record-setting experiments. In 1996, the team demonstrated atmospheric quantum-key transmission at night, quickly followed by a record-setting 0.5-kilometer point-to-point transmission in full daylight, then a 1.6-, and finally a 10-kilometer transmission. The world record for the longest QKD distribution in fiber—48 kilometers—was also held by the Los Alamos team for many years. Several of the first demonstrations of entanglement-based QKD have also been performed at the Laboratory.

In the near future, the free-space quantum cryptography system could provide secure satellite communications—using a low-orbit satellite—between cities anywhere in the world. When deployed on a spacecraft, our system can be used to generate cryptographic keys between any two users who are anywhere on the planet and can view that spacecraft. Each user would individually generate a key with the spacecraft. The second user would then be instructed to change specific bits so that the two users’ keys would match. Because the spacecraft only needs to instruct the user which bits to change, and can do so without revealing any bit values, this is a secure key-generation methodology.

On a more philosophical note, the

challenging demands of cryptography have already produced a huge growth in research into the foundations of quantum mechanics. Fundamental concepts that were previously thought to be testable only in thought experiments have been subjected to experimental verification. Many concepts, such as entanglement, that have been almost completely neglected since the early days of quantum physics have been explored and realized. This trend will continue, and we will find out to what extent the creation and control of “mesoscopic” quantum systems, that is, the netherworld between single-particle behavior and collective-particle behavior, can be performed. This research may help elucidate the puzzling transition between the quantum and classical regime. The development of quantum technology will open up other applications of quantum physics, such as quantum-enhanced sensors and improvements to atomic clocks and satellite navigation systems. Whether or not quantum cryptography becomes a widely adopted technology, we are in for an interesting next decade. ■

Acknowledgments

The Quantum Cryptography team combines the talents of numerous scientists and engineers, including those of Kevin P. McCabe, George L. Morgan, Michael J. Pigue, Steven A. Storms, Paul A. Montano, James T. Thrasher, and especially Charles G. Peterson. The authors wish to thank Derek Derkacs for technical support. We gratefully acknowledge support for the 10-kilometer free-space experiment from the National Reconnaissance Office Director’s Innovation Initiative program, administered by Col. John Comtois and Peter Hendrickson.

Further Reading

- Bennett, C. H. 1992. Quantum Cryptography: Uncertainty in the Service of Privacy. *Science* **257** (5071): 752.
- Bennett, C. H., and G. Brassard. 1984. Quantum Cryptography: Public-Key Distribution and Coin Tossing. In *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India, 1984*. p. 175. New York: IEEE.
- Bennett, C. H., G. Brassard, C. Crépeau, and U. M. Maurer. 1995. Generalized Privacy Amplification. *IEEE Trans. Inf. Theory* **41** (6): 1915.
- Bennett, C. H., G. Brassard, and A. K. Ekert. 1992. Quantum Cryptography. *Sci. Am.* **267** (4): 50.
- Hughes, R. J., D. G.L. Morgan, and C. G. Peterson. 2000. Quantum Key Distribution over a 48-km Optical Fiber Network. *J. Mod. Opt.* **47**: 533.
- Hughes, R. J., J. E. Nordholt, D. Derkacs, and C. G. Peterson. 2002. Practical Free-Space Quantum Key distribution over 10 km in Daylight and at Night. *New J. Phys.* **4**: 43. [Online]: <http://www.njp.org>
- Hughes, R. J., W. T. Buttler, P. G. Kwiat, S. K. Lamoreaux, G. L. Morgan, J. E. Nordholt, and C. G. Peterson. 2000. Free-Space Quantum Key Distribution in Daylight. *J. Mod. Opt.* **47**: 549.
- Hughes, R. J., W. T. Buttler, P. G. Kwiat, S. K. Lamoreaux, G. L. Morgan, J. E. Nordholt, and C. G. Peterson. 2000. Quantum Cryptography for Secure Satellite Communications. In *2000 IEEE Aerospace Conference Proceedings*. p. 191. New York: IEEE.
- Hughes, R., and J. Nordholt. 1999. Quantum Cryptography Takes to the Air. *Phys. World* **12** (5): 31.
- Hughes, R. J., W. T. Buttler, P. G. Kwiat, S. K. Lamoreaux, G. G. Luther, G. L. Morgan, J. E. Nordholt, and C. G. Peterson. 1999. Quantum Cryptography for Secure Free-Space Communications. *Proc. SPIE-Int. Soc. Opt. Eng.* **3615**: 98.
- Nordholt, J. E., R. J. Hughes, G. L. Morgan, C. G. Peterson, and C. C. Wipf. 2002. Present and Future Free-Space Quantum Key Distribution. *Proc. SPIE-Int. Soc. Opt. Eng.* **4635**:116.
- Schneier, B. 1995. Applied Cryptography: Protocols, Algorithms Source Code in C. New York: John Wiley & Sons.
- Singh, S. 1999. *The Code Book: The Evolution of Secrecy from Mary, Queen of Scots to Quantum Cryptography*. New York:

Jane E. (Beth) Nordholt has broad experience in quantum-key distribution, experimental astrophysics, high-energy physics, computing, and space plasma physics. Currently a technical



project leader at Los Alamos National Laboratory, she is the coinventor for the free-space quantum key distribution project and holds several patents on quantum-key distribution and spacecraft instrumentation. Beth received

four NASA group achievement awards and two Los Alamos Distinguished Performance Awards. In 2001, she received an R&D 100 Award for her work on free-space quantum cryptography from the *Research and Development* magazine. Her interests include quantum cryptography, quantum communications, quantum metrology, the composition of planetary magnetospheres, planetary science, and advanced instrumentation.

Richard J. Hughes is a Laboratory Fellow and Quantum Information Science team leader in the Neutron Science and Technology Group of the Physics Division at Los Alamos National Laboratory. He is the principal investigator for several projects in quantum computation and quantum cryptography. Richard obtained his Ph.D. in theoretical elementary particle physics from the University of Liverpool and held research positions at Oxford University and The Queen’s College, Oxford; California Institute of Technology; and CERN, the European Center for Nuclear Research. He was a distinguished visiting scientist at Oxford University and the University of Oslo. Richard was awarded the Los Alamos Fellows Prize for his work on quantum information science; he was twice awarded Los Alamos Distinguished Performance Awards for his quantum cryptography research; and he was cowinner of an R&D 100 Award for the entry “Free-Space Quantum Cryptography.” He became a Fellow of the American Physical Society in 1999. He has authored over 100 scientific papers on quantum field theory, the foundations of quantum mechanics, quantum cryptography, and quantum computation. In his spare time, Richard enjoys ultramarathon trail running over distances of up to 100 miles.



A New Face for Cryptography

Doubleday.

Wootters, W. K., and W. H. Zurek. 1982.

A Single Quantum Cannot be Cloned.

Nature **299**: 802.



Decoherence and the Transition from Quantum to Classical—*Revisited*

Wojciech H. Zurek

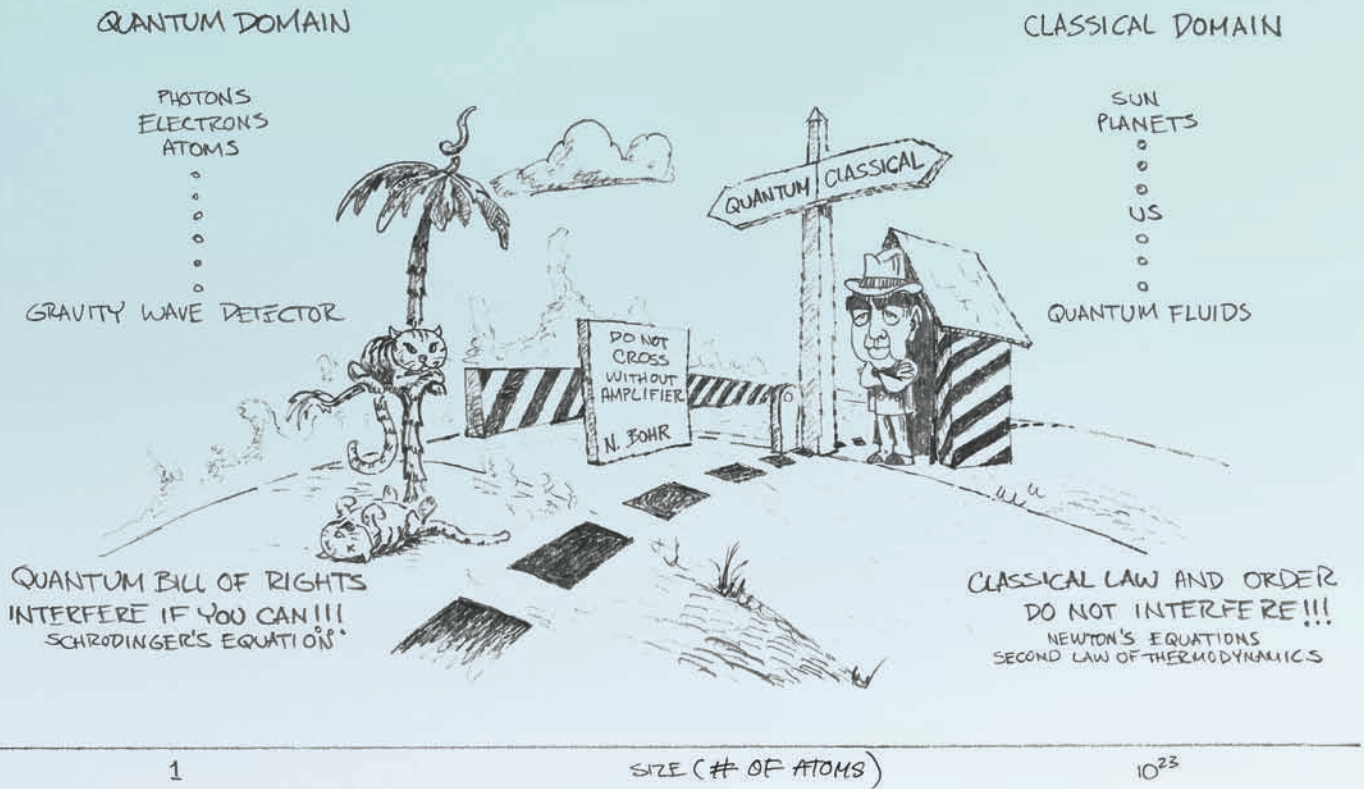
This paper has a somewhat unusual origin and, as a consequence, an unusual structure. It is built on the principle embraced by families who outgrow their dwellings and decide to add a few rooms to their existing structures instead of starting from scratch. These additions usually “show,” but the whole can still be quite pleasing to the eye, combining the old and the new in a functional way.

What follows is such a “remodeling” of the paper I wrote a dozen years ago for *Physics Today* (1991). The old text (with some modifications) is interwoven with the new text, but the additions are set off in boxes throughout this article and serve as a commentary on new developments as they relate to the original. The references appear together at the end.

In 1991, the study of decoherence was still a rather new subject, but already at that time, I had developed a feeling that most implications about the system’s “immersion” in the environment had been discovered in the preceding 10 years, so a review was in order. While writing it, I had, however, come to suspect that the small gaps in the landscape of the border territory between the quantum and the classical were actually not that small after all and that they presented excellent opportunities for further advances.

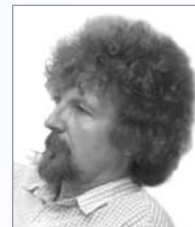
Indeed, I am surprised and gratified by how much the field has evolved over the last decade. The role of decoherence was recognized by a wide spectrum of practic-

THE BORDER TERRITORY



ing physicists as well as, beyond physics proper, by material scientists and philosophers. The study of the predictability sieve, investigations of the interface between chaotic dynamics and decoherence, and most recently, the tantalizing glimpses of the information-theoretic nature of the quantum have elucidated our understanding of the Universe. During this period, Los Alamos has grown into a leading center for the study of decoherence and related issues through the enthusiastic participation of a superb group of staff members, postdoctoral fellows, long-term visitors, and students, many of whom have become long-term collaborators. This group includes, in chronological order, Andy Albrecht, Juan Pablo Paz, Bill Wootters, Raymond Laflamme, Salman Habib, Jim Anglin, Chris Jarzynski, Kosuke Shizume, Ben Schumacher, Manny Knill, Jacek Dziarmaga, Diego Dalvit, Zbig Karkuszewski, Harold Ollivier, Roberto Onofrio, Robin Blume-Kohut, David Poulin, Lorenza Viola, and David Wallace.

Finally, I have some advice for the reader. I believe this paper should be read twice: first, just the old text alone; then—and only then—on the second reading, the whole thing. I would also recommend to the curious reader two other overviews: the draft of my *Reviews of Modern Physics* paper (Zurek 2001a) and Les Houches Lectures coauthored with Paz (Paz and Zurek 2001).



Introduction

Quantum mechanics works exceedingly well in all practical applications. No example of conflict between its predictions and experiment is known. Without quantum physics, we could not explain the behavior of the solids, the structure and function of DNA, the color of the stars, the action of lasers, or the properties of superfluids. Yet nearly a century after its inception, the debate about the relation of quantum physics to the familiar physical world continues. Why is a theory that seems to account with precision for everything we can measure still deemed lacking?

The only “failure” of quantum theory is its inability to provide a natural framework for our prejudices about the workings of the Universe. States of quantum systems evolve according to the deterministic, linear Schrödinger equation

$$i\hbar \frac{d}{dt}|\psi\rangle = H|\psi\rangle . \quad (1)$$

That is, just as in classical mechanics, given the initial state of the system and its Hamiltonian H , one can, at least in principle, compute the state at an arbitrary time. This deterministic evolution of $|\psi\rangle$ has been verified in carefully controlled experiments. Moreover, there is no indication of a border between quantum and classical at which Equation (1) would fail (see cartoon on the opener to this article).

There is, however, a very poorly controlled experiment with results so tangible and immediate that it has enormous power to convince: Our perceptions are often difficult to reconcile with the predictions of Equation (1). Why? Given almost any initial condition, the Universe described by $|\psi\rangle$ evolves into a state containing many alternatives that are never seen to coexist in our world. Moreover, while the ultimate evidence for the choice of one alternative resides in our elusive “consciousness,” there is every indication that the choice occurs much before consciousness ever gets involved and that, once made, the choice is irrevocable. Thus, at the root of our unease with quantum theory is the clash between the principle of superposition—the basic tenet of the theory reflected in the linearity of Equation (1)—and everyday classical reality in which this principle appears to be violated.

The problem of measurement has a long and fascinating history. The first widely accepted explanation of how a single outcome emerges from the multitude of potentialities was the Copenhagen Interpretation proposed by Niels Bohr (1928), who insisted that a classical apparatus is necessary to carry out measurements. Thus, quantum theory was not to be universal. The key feature of the Copenhagen Interpretation is the dividing line between quantum and classical. Bohr emphasized that the border must be mobile so that even the “ultimate apparatus”—the human nervous system—could in principle be measured and analyzed as a quantum object, provided that a suitable classical device could be found to carry out the task.

In the absence of a crisp criterion to distinguish between quantum and classical, an identification of the classical with the macroscopic has often been tentatively accepted. The inadequacy of this approach has become apparent as a result of relatively recent developments: A cryogenic version of the Weber bar—a gravity-wave detector—must be treated as a quantum harmonic oscillator even though it may weigh a ton (Braginsky et al. 1980, Caves et al. 1980). Nonclassical squeezed states can describe oscillations of suitably prepared electromagnetic fields with macroscopic numbers of photons (Teich and Saleh 1990). Finally, quantum states associated with the currents of superconducting Josephson junctions involve macroscopic numbers of electrons, but still they can tunnel between the minima of the effective potential corresponding to the opposite sense of rotation (Leggett et al. 1987, Caldeira and Leggett 1983a, Tesche 1986).

If macroscopic systems cannot be always safely placed on the classical side of the boundary, then might there be no boundary at all? The Many Worlds Interpretation (or more accurately, the Many Universes Interpretation), developed by Hugh Everett III with encouragement from John Archibald Wheeler in the 1950s, claims to do away with the boundary (Everett 1957, Wheeler 1957). In this interpretation, the entire universe is described by quantum theory. Superpositions evolve forever according to the Schrödinger equation. Each time a suitable interaction takes place between any two quantum systems, the wave function of the universe splits, developing ever more “branches.”

Initially, Everett’s work went almost unnoticed. It was taken out of mothballs over a decade later by Bryce DeWitt (1970) and DeWitt and Neill Graham (1973), who managed to upgrade its status from “virtually unknown” to “very controversial.” The Many Worlds Interpretation is a natural choice for quantum cosmology, which describes the whole Universe by means of a state vector. There is nothing more macroscopic than the Universe. It can have no a priori classical subsystems. There can be no observer “on the outside.” In this universal setting, classicality must be an emergent property of the selected observables or systems.

At first glance, the Many Worlds and Copenhagen Interpretations have little in common. The Copenhagen Interpretation demands an a priori “classical domain” with a border that enforces a classical “embargo” by letting through just one potential outcome. The Many Worlds Interpretation aims to abolish the need for the border altogether. Every potential outcome is accommodated by the ever-proliferating branches of the wave function of the Universe. The similarity between the difficulties faced by these two viewpoints becomes apparent, nevertheless, when we ask the obvious question, “Why do I, the observer, perceive only one of the outcomes?” Quantum theory, with its freedom to rotate bases in Hilbert space, does not even clearly define which states of the Universe correspond to the “branches.” Yet, our perception of a reality with alternatives—not a coherent superposition of alternatives—demands an explanation of when, where, and how it is decided what the observer actually records. Considered in this context, the Many Worlds Interpretation in its original version does not really abolish the border but pushes it all the way to the boundary between the physical Universe and consciousness. Needless to say, this is a very uncomfortable place to do physics.

In spite of the profound nature of the difficulties, recent years have seen a growing consensus that progress is being made in dealing with the measurement problem, which is the usual euphemism for the collection of interpretational conundrums described above. The key (and uncontroversial) fact has been known almost since the inception of quantum theory, but its significance for the transition from quantum to classical is being recognized only now: Macroscopic systems are never isolated from their environments. Therefore—as H. Dieter Zeh emphasized (1970)—they should not be expected to follow Schrödinger’s equation, which is applicable only to a closed system. As a result, systems usually regarded as classical suffer (or benefit) from the natural loss of quantum coherence, which “leaks out” into the environment (Zurek 1981, 1982). The resulting “decoherence” cannot be ignored when one addresses the problem of the reduction of the quantum mechanical wave packet: Decoherence imposes, in effect, the required “embargo” on the potential outcomes by allowing the observer to maintain only the records of alternatives sanctioned by decoherence and to be aware of only one of the branches—one of the “decoherent histories” in the nomenclature of Murray Gell-Mann and James Hartle (1990) and Hartle (1991).

The aim of this paper is to explain the physics and thinking behind this approach. The reader should be warned that this writer is not a disinterested witness to this development (Wigner 1983, Joos and Zeh 1985, Haake and Walls 1986, Milburn and Holmes 1986, Albrecht 1991, Hu et al. 1992), but rather, one of the proponents. I shall, nevertheless, attempt to paint a fairly honest picture and point out the difficulties as well as the accomplishments.

Decoherence in Quantum Information Processing

Much of what was written in the introduction remains valid today. One important development is the increase in experimental evidence for the validity of the quantum principle of superposition in various contexts including spectacular double-slit experiments that demonstrate interference of fullerenes (Arndt et al. 1999), the study of superpositions in Josephson junctions (Mooij et al. 1999, Friedman et al. 2000), and the implementation of Schrödinger “kittens” in atom interferometry (Chapman et al. 1995, Pfau et al. 1994), ion traps (Monroe et al. 1996) and microwave cavities (Brune et al. 1996).

In addition to confirming the superposition principle and other exotic aspects of quantum theory (such as entanglement) in novel settings, these experiments allow—as we shall see later—for a controlled investigation of decoherence.

The other important change that influenced the perception of the quantum-to-classical “border territory” is the explosion of interest in quantum information and computation. Although quantum computers were already being discussed in the 1980s, the nature of the interest has changed since Peter Shor invented his factoring algorithm. Impressive theoretical advances, including the discovery of quantum error correction and resilient quantum computation, quickly followed, accompanied by increasingly bold experimental forays. The superposition principle, once the cause of trouble for the interpretation of quantum theory, has become the central article of faith in the emerging

science of quantum information processing. This last development is discussed elsewhere in this issue, so I shall not dwell on it here.

The application of quantum physics to information processing has also transformed the nature of interest in the process of decoherence: At the time of my original review (1991), decoherence was a solution to the interpretation problem—a mechanism to impose an effective classicality on de facto quantum systems. In quantum information processing, decoherence plays two roles. Above all, it is a threat to the quantumness of quantum information. It invalidates the quantum superposition principle and thus turns quantum computers into (at best) classical computers, negating the potential power offered by the quantumness of the algorithms. But decoherence is also a necessary (although often taken for granted) ingredient in quantum information processing, which must, after all, end in a “measurement.”

The role of a measurement is to convert quantum states and quantum correlations (with their characteristic indefiniteness and malleability) into classical, definite outcomes. Decoherence leads to the environment-induced superselection (einselection) that justifies the existence of the preferred pointer states. It enables one to draw an effective border between the quantum and the classical in straightforward terms, which do not appeal to the “collapse of the wave packet” or any other such *deus ex machina*.

Correlations and Measurements

A convenient starting point for the discussion of the measurement problem and, more generally, of the emergence of classical behavior from quantum dynamics is the analysis of quantum measurements due to John von Neumann (1932). In contrast to Bohr, who assumed at the outset that the apparatus must be classical (thereby forfeiting the claim of quantum theory to universal validity), von Neumann analyzed the case of a quantum apparatus. I shall reproduce his analysis for the simplest case: a measurement on a two-state system \mathcal{S} (which can be thought of as an atom with spin 1/2) in which a quantum two-state (one bit) detector records the result.

The Hilbert space $\mathcal{H}_{\mathcal{S}}$ of the system is spanned by the orthonormal states $|\uparrow\rangle$ and $|\downarrow\rangle$, while the states $|d_{\uparrow}\rangle$ and $|d_{\downarrow}\rangle$ span the $\mathcal{H}_{\mathcal{D}}$ of the detector. A two-dimensional $\mathcal{H}_{\mathcal{D}}$ is the absolute minimum needed to record the possible outcomes. One can devise a quantum

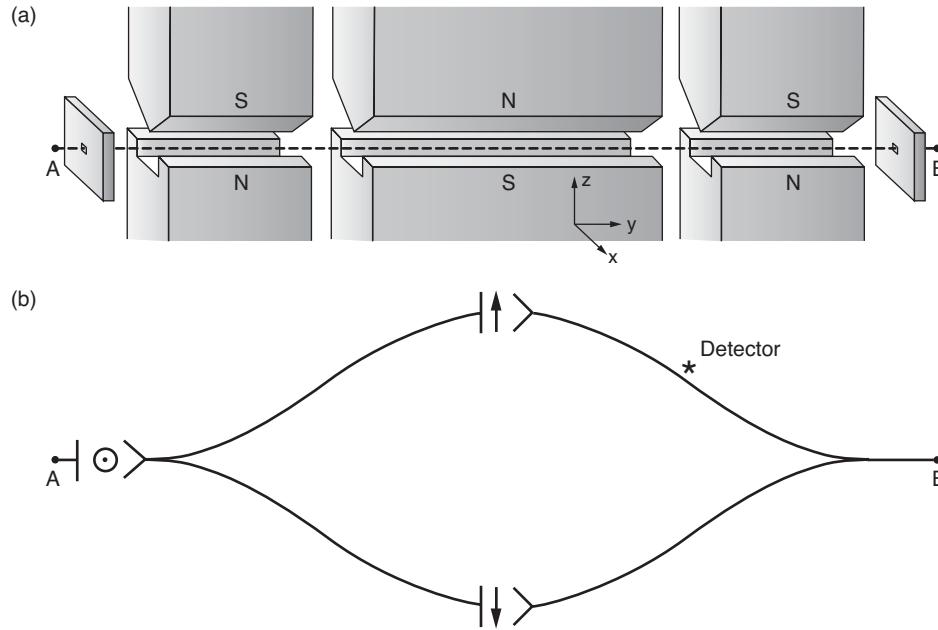


Figure 1. A Reversible Stern-Gerlach Apparatus
 The “gedanken” reversible Stern-Gerlach apparatus in (a) splits a beam of atoms into two branches that are correlated with the component of the spin of the atoms (b) and then recombines the branches before the atoms leave the device. Eugene Wigner (1963) used this gedanken experiment to show that a correlation between the spin and the location of an atom can be reversibly undone. The introduction of a one-bit (two-state) quantum detector that changes its state when the atom passes nearby prevents the reversal: The detector inherits the correlation between the spin and the trajectory, so the Stern-Gerlach apparatus can no longer undo the correlation. (This illustration was adapted with permission from Zurek 1981.)

detector (see Figure 1) that “clicks” only when the spin is in the state $|\uparrow\rangle$, that is,

$$|\uparrow\rangle |d_{\downarrow}\rangle \rightarrow |\uparrow\rangle |d_{\uparrow}\rangle, \quad (2)$$

and remains unperturbed otherwise.

I shall assume that, before the interaction, the system was in a pure state $|\psi_S\rangle$ given by

$$|\psi_S\rangle = \alpha|\uparrow\rangle + \beta|\downarrow\rangle, \quad (3)$$

with the complex coefficients satisfying $|\alpha|^2 + |\beta|^2 = 1$. The composite system starts as

$$|\Phi^i\rangle = |\psi_S\rangle |d_{\downarrow}\rangle. \quad (4)$$

Interaction results in the evolution of $|\Phi^i\rangle$ into a correlated state $|\Phi^c\rangle$:

$$|\Phi^i\rangle = (\alpha|\uparrow\rangle + \beta|\downarrow\rangle)|d_{\downarrow}\rangle \Rightarrow \alpha|\uparrow\rangle|d_{\uparrow}\rangle + \beta|\downarrow\rangle|d_{\downarrow}\rangle = |\Phi^c\rangle. \quad (5)$$

This essential and uncontroversial first stage of the measurement process can be accomplished by means of a Schrödinger equation with an appropriate interaction. It might be tempting to halt the discussion of measurements with Equation (5). After all, the correlated state vector $|\Phi^c\rangle$ implies that, if the detector is seen in the state $|d_{\uparrow}\rangle$, the system is guaranteed to be found in the state $|\uparrow\rangle$. Why ask for anything more?

The reason for dissatisfaction with $|\Phi^c\rangle$ as a description of a completed measurement is simple and fundamental: In the real world, even when we do not know the outcome of a measurement, we do know the possible alternatives, and we can safely act as if only one of those alternatives has occurred. As we shall see in the next section, such an assumption is not only unsafe but also simply wrong for a system described by $|\Phi^c\rangle$.

How then can an observer (who has not yet consulted the detector) express his ignorance about the outcome without giving up his certainty about the “menu” of the

possibilities? Quantum theory provides the right formal tool for the occasion: A density matrix can be used to describe the probability distribution over the alternative outcomes.

Von Neumann was well aware of these difficulties. Indeed, he postulated (1932) that, in addition to the unitary evolution given by Equation (1), there should be an ad hoc “process 1”—a nonunitary reduction of the state vector—that would take the pure, correlated state $|\Phi^c\rangle$ into an appropriate mixture: This process makes the outcomes independent of one another by taking the pure-state density matrix:

$$\begin{aligned} \rho^c = |\Phi^c\rangle\langle\Phi^c| = & |\alpha|^2 |\uparrow\rangle\langle\uparrow| |d_\uparrow\rangle\langle d_\uparrow| + \alpha\beta^* |\uparrow\rangle\langle\downarrow| |d_\uparrow\rangle\langle d_\downarrow| \\ & + \alpha^*\beta |\downarrow\rangle\langle\uparrow| |d_\downarrow\rangle\langle d_\uparrow| + |\beta|^2 |\downarrow\rangle\langle\downarrow| |d_\downarrow\rangle\langle d_\downarrow| , \end{aligned} \quad (6)$$

and canceling the off-diagonal terms that express purely quantum correlations (entanglement) so that the reduced density matrix with only classical correlations emerges:

$$\rho^r = |\alpha|^2 |\uparrow\rangle\langle\uparrow| |d_\uparrow\rangle\langle d_\uparrow| + |\beta|^2 |\downarrow\rangle\langle\downarrow| |d_\downarrow\rangle\langle d_\downarrow| . \quad (7)$$

Why is the reduced ρ^r easier to interpret as a description of a completed measurement than ρ^c ? After all, both ρ^r and ρ^c contain identical diagonal elements. Therefore, both outcomes are still potentially present. So what—if anything—was gained at the substantial price of introducing a nonunitary process 1?

The Question of Preferred Basis: What Was Measured?

The key advantage of ρ^r over ρ^c is that its coefficients may be interpreted as classical probabilities. The density matrix ρ^r can be used to describe the alternative states of a composite spin-detector system that has classical correlations. Von Neumann’s process 1 serves a similar purpose to Bohr’s “border” even though process 1 leaves all the alternatives in place. When the off-diagonal terms are absent, one can nevertheless safely maintain that the apparatus, as well as the system, is each separately in a definite but unknown state, and that the correlation between them still exists in the preferred basis defined by the states appearing on the diagonal. By the same token, the identities of two halves of a split coin placed in two sealed envelopes may be unknown but are classically correlated. Holding one unopened envelope, we can be sure that the half it contains is either “heads” or “tails” (and not some superposition of the two) and that the second envelope contains the matching alternative.

By contrast, it is impossible to interpret ρ^c as representing such “classical ignorance.” In particular, even the set of the alternative outcomes is not decided by ρ^c ! This circumstance can be illustrated in a dramatic fashion by choosing $\alpha = -\beta = 1/\sqrt{2}$ so that the density matrix ρ^c is a projection operator constructed from the correlated state

$$|\Phi^c\rangle = (|\uparrow\rangle|d_\uparrow\rangle - |\downarrow\rangle|d_\downarrow\rangle)/\sqrt{2} . \quad (8)$$

This state is invariant under the rotations of the basis. For instance, instead of the eigenstates of $|\uparrow\rangle$ and $|\downarrow\rangle$ of $\hat{\sigma}_z$ one can rewrite $|\Phi^c\rangle$ in terms of the eigenstates of $\hat{\sigma}_x$:

$$|\odot\rangle = (|\uparrow\rangle + |\downarrow\rangle)/\sqrt{2} , \quad (9a)$$

$$|\otimes\rangle = (|\uparrow\rangle - |\downarrow\rangle)/\sqrt{2} . \quad (9b)$$

This representation immediately yields

$$|\Phi^c\rangle = -(|\odot\rangle|d_\odot\rangle - |\otimes\rangle|d_\otimes\rangle)/\sqrt{2} , \quad (10)$$

where

$$|d_\odot\rangle = (|d_\downarrow\rangle - |d_\uparrow\rangle)/\sqrt{2} \quad \text{and} \quad |d_\otimes\rangle = (|d_\uparrow\rangle + |d_\downarrow\rangle)/\sqrt{2} \quad (11)$$

are, as a consequence of the superposition principle, perfectly “legal” states in the Hilbert space of the quantum detector. Therefore, the density matrix

$$\rho^c = |\Phi^c\rangle\langle\Phi^c|$$

could have many (in fact, infinitely many) different states of the subsystems on the diagonal.

This freedom to choose a basis should not come as a surprise. Except for the notation, the state vector $|\Phi^c\rangle$ is the same as the wave function of a pair of maximally correlated (or entangled) spin-1/2 systems in David Bohm’s version (1951) of the Einstein-Podolsky-Rosen (EPR) paradox (Einstein et al. 1935). And the experiments that show that such nonseparable quantum correlations violate Bell’s inequalities (Bell 1964) are demonstrating the following key point: The states of the two spins in a system described by $|\Phi^c\rangle$ are not just unknown, but rather they cannot exist before the “real” measurement (Aspect et al. 1981, 1982). We conclude that when a detector is quantum, a superposition of records exists and is a record of a superposition of outcomes—a very nonclassical state of affairs.

Missing Information and Decoherence

Unitary evolution condemns every closed quantum system to “purity.” Yet, if the outcomes of a measurement are to become independent events, with consequences that can be explored separately, a way must be found to dispose of the excess information. In the previous sections, quantum correlation was analyzed from the point of view of its role in acquiring information. Here, I shall discuss the flip side of the story: Quantum correlations can also disperse information throughout the degrees of freedom that are, in effect, inaccessible to the observer. Interaction with the degrees of freedom external to the system—which we shall summarily refer to as the environment—offers such a possibility.

Reduction of the state vector, $\rho^c \Rightarrow \rho^r$, decreases the information available to the observer about the composite system \mathcal{SD} . The information loss is needed if the outcomes are to become classical and thereby available as initial conditions to predict the future. The effect of this loss is to increase the entropy $\mathcal{H} = -\text{Tr}\rho \ln\rho$ by an amount

$$\Delta\mathcal{H} = \mathcal{H}(\rho^r) - \mathcal{H}(\rho^c) = -(|\alpha|^2 \ln|\alpha|^2 + |\beta|^2 \ln|\beta|^2) . \quad (12)$$

Entropy must increase because the initial state described by ρ^c was pure, $\mathcal{H}(\rho^c) = 0$, and the reduced state is mixed. Information gain—the objective of the measurement—is accomplished only when the observer interacts and becomes correlated with the detector in the already precollapsed state ρ^r .

To illustrate the process of the environment-induced decoherence, consider a system \mathcal{S} , a detector \mathcal{D} , and an environment \mathcal{E} . The environment is also a quantum system. Following the first step of the measurement process—establishment of a correlation as shown in Equation (5)—the environment similarly interacts and becomes correlated with the apparatus:

$$|\Phi^c\rangle |\mathcal{E}_0\rangle = (\alpha|\uparrow\rangle|d_\uparrow\rangle + \beta|\downarrow\rangle|d_\downarrow\rangle) |\mathcal{E}_0\rangle \Rightarrow \alpha|\uparrow\rangle|d_\uparrow\rangle |\mathcal{E}_\uparrow\rangle + \beta|\downarrow\rangle|d_\downarrow\rangle |\mathcal{E}_\downarrow\rangle = |\Psi\rangle . \quad (13)$$

The final state of the combined $\mathcal{SD}\mathcal{E}$ “von Neumann chain” of correlated systems extends the correlation beyond the \mathcal{SD} pair. When the states of the environment $|\mathcal{E}_i\rangle$ corresponding to the states $|d_\uparrow\rangle$ and $|d_\downarrow\rangle$ of the detector are orthogonal, $\langle \mathcal{E}_i | \mathcal{E}_{i'} \rangle = \delta_{ii'}$, the density matrix for the detector-system combination is obtained by ignoring (tracing over) the information in the uncontrolled (and unknown) degrees of freedom

$$\rho_{\mathcal{DS}} = \text{Tr}_{\mathcal{E}} |\Psi\rangle\langle\Psi| = \sum_i \langle \mathcal{E}_i | \Psi\rangle\langle\Psi | \mathcal{E}_i \rangle = |\alpha|^2 |\uparrow\rangle\langle\uparrow| |d_\uparrow\rangle\langle d_\uparrow| + |\beta|^2 |\downarrow\rangle\langle\downarrow| |d_\downarrow\rangle\langle d_\downarrow| = \rho^r . \quad (14)$$

The resulting ρ^r is precisely the reduced density matrix that von Neumann called for. Now, in contrast to the situation described by Equations (9)–(11), a superposition of the records of the detector states is no longer a record of a superposition of the state of the system. A preferred basis of the detector, sometimes called the “pointer basis” for obvious reasons, has emerged. Moreover, we have obtained it—or so it appears—without having to appeal to von Neumann’s nonunitary process 1 or anything else beyond the ordinary, unitary Schrödinger evolution. The preferred basis of the detector—or for that matter, of any open quantum system—is selected by the dynamics.

Not all aspects of this process are completely clear. It is, however, certain that the detector–environment interaction Hamiltonian plays a decisive role. In particular, when the interaction with the environment dominates, eigenspaces of any observable Λ that commutes with the interaction Hamiltonian,

$$[\Lambda, H_{int}] = 0 , \quad (15)$$

invariably end up on the diagonal of the reduced density matrix (Zurek 1981, 1982). This commutation relation has a simple physical implication: It guarantees that the pointer observable Λ will be a constant of motion, a conserved quantity under the evolution generated by the interaction Hamiltonian. Thus, when a system is in an eigenstate of Λ , interaction with the environment will leave it unperturbed.

In the real world, the spreading of quantum correlations is practically inevitable. For example, when in the course of measuring the state of a spin-1/2 atom (see Figure 1b), a photon had scattered from the atom while it was traveling along one of its two alternative routes, this interaction would have resulted in a correlation with the environment and would have necessarily led to a loss of quantum coherence. The density matrix of the \mathcal{SD} pair would have lost its off-diagonal terms. Moreover, given that it is impossible to catch up with the photon, such loss of coherence would have been irreversible. As we shall see later, irreversibility could also arise from more familiar, statistical causes: Environments are notorious for having large numbers of interacting degrees of freedom, making extraction of lost information as difficult as reversing trajectories in the Boltzmann gas.

Quantum Discord—A Measure of Quantumness

The contrast between the density matrices in Equations (6) and (7) is stark and obvious. In particular, the entanglement between the system and the detector in ρ^c is obviously quantum—classical systems cannot be entangled. The argument against the “ignorance” interpretation of ρ^c still stands. Yet we would like to have a quantitative measure of how much is classical (or how much is quantum) about the correlations of a state represented by a general density matrix. Such a measure of the quantumness of correlation was devised recently (Ollivier and Zurek 2002). It is known as quantum discord. Of the several closely related definitions of discord, we shall select one that is easiest to explain. It is based on mutual information—an information-theoretic measure of how much easier it is to describe the state of a pair of objects $(\mathcal{S}, \mathcal{D})$ jointly rather than separately. One formula for mutual information $I(\mathcal{S}:\mathcal{D})$ is simply

$$I(\mathcal{S}:\mathcal{D}) = \mathcal{H}(\mathcal{S}) + \mathcal{H}(\mathcal{D}) - \mathcal{H}(\mathcal{S}, \mathcal{D}),$$

where $\mathcal{H}(\mathcal{S})$ and $\mathcal{H}(\mathcal{D})$ are the entropies of \mathcal{S} and \mathcal{D} , respectively, and $\mathcal{H}(\mathcal{S}, \mathcal{D})$ is the joint entropy of the two. When \mathcal{S} and \mathcal{D} are not correlated (statistically independent),

$$\mathcal{H}(\mathcal{S}, \mathcal{D}) = \mathcal{H}(\mathcal{S}) + \mathcal{H}(\mathcal{D}),$$

and $I(\mathcal{S}:\mathcal{D}) = 0$. By contrast, when there is a perfect classical correlation between them (for example, two copies of the same book), $\mathcal{H}(\mathcal{S}, \mathcal{D}) = \mathcal{H}(\mathcal{S}) = \mathcal{H}(\mathcal{D}) = I(\mathcal{S}:\mathcal{D})$. Perfect classical correlation implies that, when we find out all about one of them, we also know everything about the other, and the conditional entropy $\mathcal{H}(\mathcal{S}|\mathcal{D})$ (a measure of the uncertainty about \mathcal{S} after the state of \mathcal{D} is found out) disappears. Indeed, classically, the joint entropy $\mathcal{H}(\mathcal{S}, \mathcal{D})$ can always be decomposed into, say, $\mathcal{H}(\mathcal{D})$, which measures the information missing about \mathcal{D} , and the conditional entropy $\mathcal{H}(\mathcal{S}|\mathcal{D})$. Information is still missing about \mathcal{S} even after the state of \mathcal{D} has been determined: $\mathcal{H}(\mathcal{S}, \mathcal{D}) = \mathcal{H}(\mathcal{D}) + \mathcal{H}(\mathcal{S}|\mathcal{D})$. This expression for the joint entropy suggests an obvious rewrite of the preceding definition of mutual information into a classically identical form, namely,

$$J(\mathcal{S}:\mathcal{D}) = \mathcal{H}(\mathcal{S}) + \mathcal{H}(\mathcal{D}) - (\mathcal{H}(\mathcal{D}) + \mathcal{H}(\mathcal{S}|\mathcal{D})).$$

Here, we have abstained from the obvious (and perfectly justified from a classical viewpoint) cancellation in order to emphasize the central feature of quan-

tumness: In quantum physics, the state collapses into one of the eigenstates of the measured observable. Hence, a state of the object is redefined by a measurement. Thus, the joint entropy can be defined in terms of the conditional entropy only after the measurement used to access, say, \mathcal{D} , has been specified. In that case,

$$\mathcal{H}_{|d_k\rangle}(\mathcal{S}, \mathcal{D}) = (\mathcal{H}(\mathcal{D}) + \mathcal{H}(\mathcal{S}|\mathcal{D}))_{|d_k\rangle}.$$

This type of joint entropy expresses the ignorance about the pair $(\mathcal{S}, \mathcal{D})$ after the observable with the eigenstates $\{|d_k\rangle\}$ has been measured on \mathcal{D} . Of course, $\mathcal{H}_{|d_k\rangle}(\mathcal{S}, \mathcal{D})$ is not the only way to define the entropy of the pair. One can also compute a basis-independent joint entropy $\mathcal{H}(\mathcal{S}, \mathcal{D})$, the von Neumann entropy of the pair. Since these two definitions of joint entropy do not coincide in the quantum case, we can define a basis-dependent quantum discord

$$\delta_{|d_k\rangle}(\mathcal{S}|\mathcal{D}) = I - J = (\mathcal{H}(\mathcal{D}) + \mathcal{H}(\mathcal{S}|\mathcal{D}))_{|d_k\rangle} - \mathcal{H}(\mathcal{S}, \mathcal{D})$$

as the measure of the extent by which the underlying density matrix describing \mathcal{S} and \mathcal{D} is perturbed by a measurement of the observable with the eigenstates $\{|d_k\rangle\}$. States of classical objects—or classical correlations—are “objective:” They exist independent of measurements. Hence, when there is a basis $\{|\hat{d}_k\rangle\}$ such that the minimum discord evaluated for this basis disappears,

$$\hat{\delta}(\mathcal{S}|\mathcal{D}) = \min_{\{|d_k\rangle\}} (\mathcal{H}(\mathcal{S}, \mathcal{D}) - (\mathcal{H}(\mathcal{D}) + \mathcal{H}(\mathcal{S}|\mathcal{D}))_{|d_k\rangle}) = 0,$$

the correlation can be regarded as effectively classical (or more precisely, as “classically accessible through \mathcal{D} ”). One can then show that there is a set of probabilities associated with the basis $\{|d_k\rangle\}$ that can be treated as classical. It is straightforward to see that, when \mathcal{S} and \mathcal{D} are entangled (for example, $\rho^c = |\phi^c\rangle\langle\phi^c|$), then $\hat{\delta} > 0$ in all bases. By contrast, if we consider ρ^r , discord disappears in the basis $\{|d_\uparrow\rangle, |d_\downarrow\rangle\}$ so that the underlying correlation is effectively classical.

It is important to emphasize that quantum discord is not just another measure of entanglement but a genuine measure of the quantumness of correlations. In situations involving measurements and decoherence, quantumness disappears for the preferred set of states that are effectively classical and thus serves as an indicator of the pointer basis, which as we shall see, emerges as a result of decoherence and einselection.

Decoherence: How Long Does It Take?

A tractable model of the environment is afforded by a collection of harmonic oscillators (Feynman and Vernon 1963, Dekker 1981, Caldeira and Leggett 1983a, 1983b, 1985, Joos and Zeh 1985, Hu et al. 1992) or, equivalently, by a quantum field (Unruh and Zurek 1989). If a particle is present, excitations of the field will scatter off the particle. The resulting “ripples” will constitute a record of its position, shape, orientation, and so on, and most important, its instantaneous location and hence its trajectory.

A boat traveling on a quiet lake or a stone that fell into water will leave such an imprint on the water surface. Our eyesight relies on the perturbation left by the objects on the preexisting state of the electromagnetic field. Hence, it is hardly surprising that an imprint is left whenever two quantum systems interact, even when “nobody is looking,” and even when the lake is stormy and full of preexisting waves, and the field is full of excitations—that is, when the environment starts in equilibrium at some finite temperature. “Messy” initial states of the environment make it difficult to decipher the record, but do not preclude its existence.

A specific example of decoherence—a particle at position x interacting with a scalar field ϕ (which can be regarded as a collection of harmonic oscillators) through the Hamiltonian

$$H_{int} = \epsilon x d\phi/dt \quad , \quad (16)$$

where ϵ is the strength of the coupling, has been extensively studied by many, including the investigators just referenced. The conclusion is easily formulated in the so-called “high-temperature limit,” in which only thermal-excitation effects of the field ϕ are taken into account and the effect of zero-point vacuum fluctuations is neglected.

In this case, the density matrix $\rho(x, x')$ of the particle in the position representation evolves according to the master equation

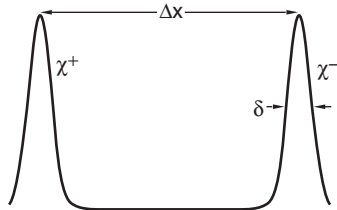
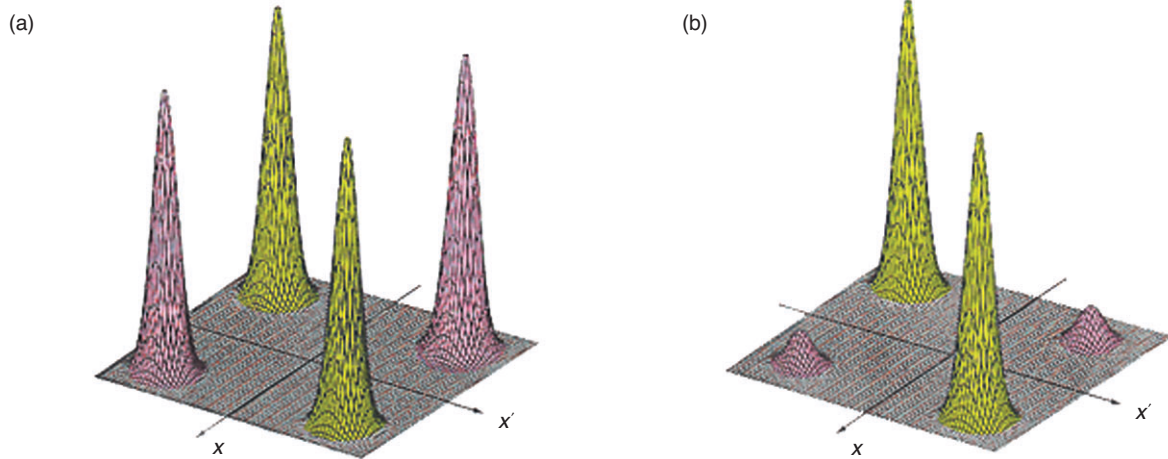


Figure 2. A “Schrödinger Cat” State or a Coherent Superposition
 This cat state $\varphi(x)$, the coherent superposition of two Gaussian wave packets of Equation (18), could describe a particle in a superposition of locations inside a Stern-Gerlach apparatus (see Figure 1) or the state that develops in the course of a double-slit experiment. The phase between the two components has been chosen to be zero.

$$\dot{\rho} = \underbrace{-\frac{i}{\hbar}[H, \rho]}_{\dot{\rho} = -\text{FORCE} = \nabla V} \quad - \quad \underbrace{\gamma(x-x')\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)}_{\dot{\rho} = -\gamma p} \rho \quad - \quad \underbrace{\frac{2m\gamma k_B T}{\hbar^2}(x-x')^2 \rho}_{\text{Classical Phase Space}} \quad , \quad (17)$$

where H is the particle’s Hamiltonian (although with the potential $V(x)$ adjusted because of H_{int}), γ is the relaxation rate, k_B is the Boltzmann constant, and T is the temperature of the field. Equation (17) is obtained by first solving exactly the Schrödinger equation for a particle plus the field and then tracing over the degrees of freedom of the field.

I will not analyze Equation (17) in detail but just point out that it naturally separates into three distinct terms, each of them responsible for a different aspect of the effectively classical behavior. The first term—the von Neumann equation (which can be derived from the Schrödinger equation)—generates reversible classical evolution of the expectation value of any observable that has a classical counterpart regardless of the form of ρ (Ehrenfest’s theorem). The second term causes dissipation. The relaxation rate $\gamma = \eta/2m$ is proportional to the viscosity $\eta = \epsilon^2/2$ because of the interaction with the scalar field. That interaction causes a decrease in the average momentum and loss of energy. The last term also has a classical counterpart: It is responsible for fluctuations or random “kicks” that lead to Brownian motion. We shall see this in more detail in the next section.



For our purposes, the effect of the last term on quantum superpositions is of greatest interest. I shall show that it destroys quantum coherence, eliminating off-diagonal terms responsible for quantum correlations between spatially separated pieces of the wave packet. It is therefore responsible for the classical structure of the phase space, as it converts superpositions into mixtures of localized wave packets which, in the classical limit, turn into the familiar points in phase space. This effect is best illustrated by an example. Consider the “cat” state shown in Figure 2, where the wave function of a particle is given by a coherent superposition of two Gaussians: $\varphi(x) = (\chi^+(x) + \chi^-(x))/2^{1/2}$ and the Gaussians are

$$\chi^\pm(x) = \langle x | \pm \rangle \sim \exp \left[-\frac{\left(x \pm \frac{\Delta x}{2}\right)^2}{4\delta^2} \right]. \quad (18)$$

For the case of wide separation ($\Delta x \gg \delta$), the corresponding density matrix $\rho(x, x') = \varphi(x) \varphi^*(x')$ has four peaks: Two on the diagonal defined by $x = x'$, and two off the diagonal for which x and x' are very different (see Figure 3). Quantum coherence is due to the off-diagonal peaks. As those peaks disappear, position emerges as an approximate preferred basis.

The last term of Equation (17), which is proportional to $(x - x')^2$, has little effect on the diagonal peaks. By contrast, it has a large effect on the off-diagonal peaks for which $(x - x')^2$ is approximately the square of the separation $(\Delta x)^2$. In particular, it causes the

off-diagonal peaks to decay at the rate $\frac{d}{dt}(\rho^{+-}) \sim 2\gamma m k_B T / \hbar^2 (\Delta x)^2 \rho^{+-} = \tau_D^{-1} \rho^{+-}$.

It follows that quantum coherence will disappear on a decoherence time scale (Zurek 1984):

$$\tau_D \cong \gamma^{-1} \left(\frac{\lambda_{dB}}{\Delta x} \right)^2 = \tau_R \left(\frac{\hbar}{\Delta x \sqrt{2m k_B T}} \right)^2, \quad (19)$$

where $\lambda_{dB} = \hbar/(2m k_B T)^{1/2}$ is the thermal de Broglie wavelength. For macroscopic objects, the decoherence time τ_D is typically much less than the relaxation time $\tau_R = \gamma^{-1}$.

Figure 3. Evolution of the Density Matrix for the Schrödinger Cat State in Figure 2

(a) This plot shows the density matrix for the cat state in Figure 2 in the position representation $\rho(x, x') = \varphi(x)\varphi^*(x')$. The peaks near the diagonal (green) correspond to the two possible locations of the particle. The peaks away from the diagonal (red) are due to quantum coherence. Their existence and size demonstrate that the particle is not in either of the two approximate locations but in a coherent superposition of them. (b) Environment-induced decoherence causes decay of the off-diagonal terms of $\rho(x, x')$. Here, the density matrix in (a) has partially decohered. Further decoherence would result in a density matrix with diagonal peaks only. It can then be regarded as a classical probability distribution with an equal probability of finding the particle in either of the locations corresponding to the Gaussian wave packets.

For a system at temperature $T = 300$ kelvins with mass $m = 1$ gram and separation $\Delta x = 1$ centimeter, the ratio of the two time scales is $\tau_D/\tau_R \sim 10^{-40}$! Thus, even if the relaxation rate were of the order of the age of the Universe, $\sim 10^{17}$ seconds, quantum coherence would be destroyed in $\tau_D \sim 10^{-23}$ second.

For microscopic systems and, occasionally, even for very macroscopic ones, the decoherence times are relatively long. For an electron ($m_e = 10^{-27}$ grams), τ_D can be much larger than the other relevant time scales on atomic and larger energy and distance scales. For a massive Weber bar, tiny Δx ($\sim 10^{-17}$ centimeter) and cryogenic temperatures suppress decoherence. Nevertheless, the macroscopic nature of the object is certainly crucial in facilitating the transition from quantum to classical.

Experiments on Decoherence

A great deal of work on master equations and their derivations in different situations has been conducted since 1991, but in effect, most of the results described above stand. A summary can be found in Paz and Zurek (2001) and a discussion of the caveats to the simple conclusions regarding decoherence rates appears in James Anglin et al. (1997).

Perhaps the most important development in the study of decoherence is on the experimental front. In the past decade, several experiments probing decoherence in various systems have been carried out. In particular, Michel Brune, Serge Haroche, Jean-Michel Raimond, and their colleagues at École Normale Supérieure in Paris (Brune et al. 1996, Haroche 1998) have performed a series of microwave cavity experiments in which they manipulate electromagnetic fields into a Schrödinger-cat-like superposition using rubidium atoms. They probe the ensuing loss of quantum coherence. These experiments have confirmed the basic tenets of decoherence theory. Since then, the French scientists have applied the same techniques to implement various quantum information-processing ventures. They are in the process of upgrading their equipment in order to produce “bigger and better” Schrödinger cats and to study their decoherence.

A little later, Wineland, Monroe, and coworkers (Turchette et al. 2000) used ion traps (set up to implement a fragment of one of the quantum computer designs) to study the decoherence of ions due to radiation. Again, theory was confirmed, further advancing the status of decoherence as both a key ingredient of the explanation of the emergent classicality and a threat to quantum computation. In addition to these developments, which test various aspects of decoherence induced by a real or simulated “large environment,” Pritchard and his coworkers at the Massachusetts Institute of Technology have carried out a beautiful sequence of experiments by using atomic interferometry in order to investigate the role of information transfer between atoms and photons (see Kokorowski et al. 2001 and other references therein). Finally, “analogue experiments” simulating the behavior of the Schrödinger equation in optics (Cheng and Raymer 1999) have explored some of the otherwise difficult-to-access corners of the parameter space.

In addition to these essentially mesoscopic Schrödinger-cat decoherence experiments, designs of much more substantial “cats” (for example, mirrors in superpositions of quantum states) are being investigated in several laboratories.

Classical Limit of Quantum Dynamics

The Schrödinger equation was deduced from classical mechanics in the Hamilton-Jacobi form. Thus, it is no surprise that it yields classical equations of motion when \hbar can be regarded as small. This fact, along with Ehrenfest's theorem, Bohr's correspondence principle, and the kinship of quantum commutators with the classical Poisson brackets, is part of the standard lore found in textbooks. However, establishing the quantum-classical correspondence involves the states as well as the equations of motion. Quantum mechanics is formulated in Hilbert space, which can accommodate localized wave packets with sensible classical limits as well as the most bizarre superpositions. By contrast, classical dynamics happens in phase space.

To facilitate the study of the transition from quantum to classical behavior, it is convenient to employ the Wigner transform of a wave function $\psi(x)$:

$$W(x, p) = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} e^{ipy/\hbar} \psi^* \left(x + \frac{y}{2} \right) \psi \left(x - \frac{y}{2} \right) dy, \quad (20)$$

which expresses quantum states as functions of position and momentum.

The Wigner distribution $W(x, p)$ is real, but it can be negative. Hence, it cannot be regarded as a probability distribution. Nevertheless, when integrated over one of the two variables, it yields the probability distribution for the other (for example, $\int W(x, p) dp = |\psi(x)|^2$). For a minimum uncertainty wave packet, $\psi(x) = \pi^{-1/4} \delta^{-1/2} \exp\{- (x - x_0)^2 / 2\delta^2 + ip_0 x / \hbar\}$, the Wigner distribution is a Gaussian in both x and p :

$$W(x, p) = \frac{1}{\pi\hbar} \exp \left\{ - \frac{(x - x_0)^2}{\delta^2} - \frac{(p - p_0)^2 \delta^2}{\hbar^2} \right\}. \quad (21)$$

It describes a system that is localized in both x and p . Nothing else that Hilbert space has to offer is closer to approximating a point in classical phase space. The Wigner distribution is easily generalized to the case of a general density matrix $\rho(x, x')$:

$$W(x, p) = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} e^{ipy/\hbar} \rho \left(x - \frac{y}{2}, x + \frac{y}{2} \right) dy, \quad (22)$$

where $\rho(x, x')$ is, for example, the reduced density matrix of the particle discussed before.

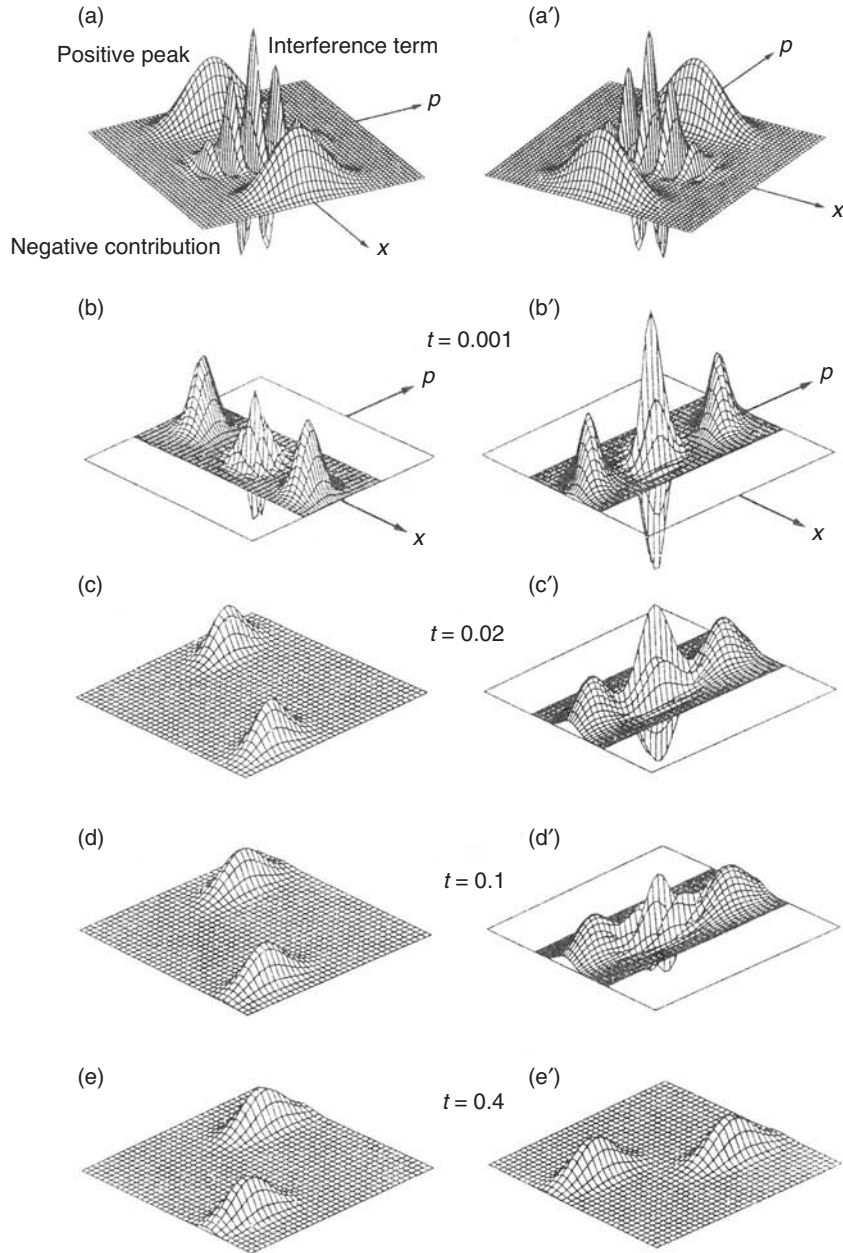
The phase-space nature of the Wigner transform suggests a strategy for exhibiting classical behavior: Whenever $W(x, p)$ represents a mixture of localized wave packets—as in Equation (21)—it can be regarded as a classical probability distribution in the phase space. However, when the underlying state is truly quantum, as is the superposition in Figure 2, the corresponding Wigner distribution function will have alternating sign—see Figure 4(a). This property alone will make it impossible to regard the function as a probability distribution in phase space. The Wigner function in Figure 4(a) is

$$W(x, p) \sim \frac{(W^+ + W^-)}{2} + \frac{1}{\pi\hbar} \exp \left\{ - \frac{p^2 \delta^2}{\hbar^2} - \frac{x^2}{\delta^2} \right\} \cdot \cos \left(\frac{\Delta x}{\hbar} p \right), \quad (23)$$

where the Gaussians W^+ and W^- are Wigner transforms of the Gaussian wave packets χ^+ and χ^- . If the underlying state had been a mixture of χ^+ and χ^- rather than a superposition, the Wigner function would have been described by the same two Gaussians W^+ and W^- , but the oscillating term would have been absent.

Figure 4. Wigner Distributions and Their Decoherence for Coherent Superpositions

(a) The Wigner distribution $W(x, p)$ is plotted as a function of x and p for the cat state of Figure 2. Note the two separate positive peaks as well as the oscillating interference term in between them. This distribution cannot be regarded as a classical probability distribution in phase space because it has negative contributions. (b–e) Decoherence produces diffusion in the direction of the momentum. As a result, the negative and positive ripples of the interference term in $W(x, p)$ diffuse into each other and cancel out. This process is almost instantaneous for open macroscopic systems. In the appropriate limit, the Wigner function has a classical structure in phase space and evolves in accord with the equations of classical dynamics. (a'–e') The analogous initial Wigner distribution and its evolution for a superposition of momenta are shown. The interference terms disappear more slowly on a time scale dictated by the dynamics of the system: Decoherence is caused by the environment coupling to (that is, monitoring) the position of the system—see Equation(16). So, for a superposition of momenta, it will start only after different velocities move the two peaks into different locations.



The equation of motion for $W(x, p)$ of a particle coupled to an environment can be obtained from Equation (17) for $\rho(x, x')$:

$$\frac{\partial W}{\partial t} = \underbrace{-\frac{p}{m} \frac{\partial}{\partial x} W + \frac{\partial V}{\partial x} \frac{\partial}{\partial p} W}_{\text{Liouville Equation}} + \underbrace{2\gamma \frac{\partial}{\partial p} p W}_{\text{Friction}} + \underbrace{D \frac{\partial^2 W}{\partial p^2}}_{\text{Decoherence}}, \quad (24)$$

where V is the renormalized potential and $D = 2m\gamma k_B T = \eta k_B T$. The three terms of this equation correspond to the three terms of Equation (17).

The first term is easily identified as a classical Poisson bracket $\{H, W\}$. That is,

The Predictability Sieve

Since 1991, understanding the emergence of preferred pointer states during the process of decoherence has advanced a great deal. Perhaps the most important advance is the predictability sieve (Zurek 1993, Zurek et al. 1993), a more general definition of pointer states that applies even when the interaction with the environment does not dominate over the self-Hamiltonian of the system. The predictability sieve sifts through the Hilbert space of a system interacting with its environment and selects states that are most predictable. Motivation for the predictability sieve comes from the observation that classical states exist or evolve predictably. Therefore, selecting quantum states that retain predictability in spite of the coupling to the environment is the obvious strategy in search of classicality. To implement the predictability sieve, we imagine a (continuously infinite) list of all the pure states $\{|\psi\rangle\}$ in the Hilbert space of the system in question. Each of them would evolve, after a time t , into a density matrix $\rho_{|\psi\rangle}(t)$. If the system were isolated, all the density matrices would have the form $\rho_{|\psi\rangle}(t) = |\psi(t)\rangle\langle\psi(t)|$ of projection operators, where $|\psi(t)\rangle$ is the appropriate solution of the Schrödinger equation. But when the system is coupled to the environment (that is, the system is “open”), $\rho_{|\psi\rangle}(t)$ is truly mixed and has a nonzero von Neumann entropy. Thus, one can compute $\mathcal{H}(\rho_{|\psi\rangle}(t)) = -\text{Tr}\rho_{|\psi\rangle}(t) \log\rho_{|\psi\rangle}(t)$, thereby defining a functional on the Hilbert space \mathcal{H}_S of the system, $|\psi\rangle \rightarrow \mathcal{H}(|\psi\rangle, t)$.

An obvious way to look for predictable, effectively classical states is to seek a subset of all $\{|\psi\rangle\}$ that minimize $\mathcal{H}(|\psi\rangle, t)$ after a certain, sufficiently long time t . When such preferred pointer states exist, are well defined (that is, the minimum of the entropy $\mathcal{H}(|\psi\rangle, t)$ differs significantly for pointer states from the average value), and are reasonably stable (that is, after the initial decoherence

time, the set of preferred states is reasonably insensitive to the precise value of t), one can consider them as good candidates for the classical domain. Figure A illustrates an implementation of the predictability sieve strategy using a different, simpler measure of predictability—purity ($\text{Tr}\rho^2$)—rather than the von Neumann entropy, which is much more difficult to compute.

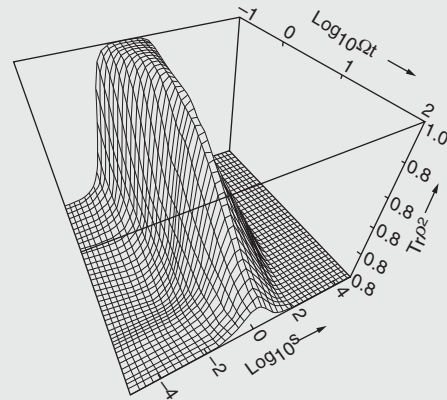


Figure A. The Predictability Sieve for the Underdamped Harmonic Oscillator

One measure of predictability is the so-called purity $\text{Tr}\rho^2$, which is plotted as a function of time for mixtures of minimum uncertainty wave packets in an underdamped harmonic oscillator with $\gamma/\omega = 10^{-4}$. The wave packets start with different squeeze parameters s . $\text{Tr}\rho^2$ serves as a measure of the purity of the reduced density matrix ρ . The predictability sieve favors coherent states ($s = 1$), which have the shape of a ground state, that is, the same spread in position and momentum when measured in units natural for the harmonic oscillator. Because they are the most predictable (more than the energy eigenstates), they are expected to play the crucial role of the pointer basis in the transition from quantum to classical.

if $w(x, p)$ is a familiar classical probability density in phase space, then it evolves according to:

$$\frac{\partial w}{\partial t} = -\frac{\partial w}{\partial x} \frac{\partial H}{\partial p} + \frac{\partial w}{\partial p} \frac{\partial H}{\partial x} = \{H, w\} = Lw \quad (25)$$

where L stands for the Liouville operator. Thus, classical dynamics in its Liouville form follows from quantum dynamics at least for the harmonic oscillator case, which is described rigorously by Equations (17) and (24). (For more general $V(x)$, the Poisson bracket would have to be supplemented by quantum corrections of order \hbar .) The second term of Equation (24) represents friction. The last term results in the diffusion of $W(x, p)$ in momentum at the rate given by D .

Continued on page 104

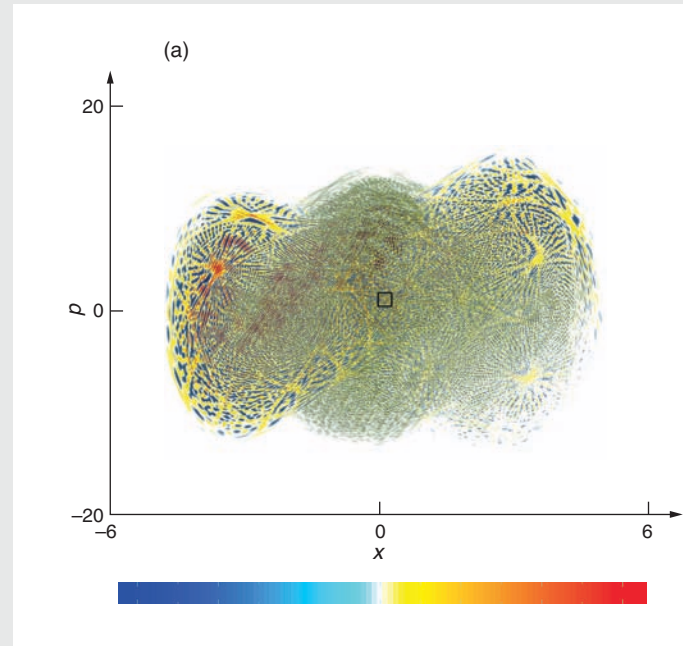
Quantum Chaos and Phase-Space Aspects of Quantum-Classical Correspondence

Classical mechanics “happens” in phase space. It is therefore critically important to show that quantum theory can—in the presence of decoherence—reproduce the basic structure of classical phase space and that it can emulate classical dynamics. The argument put forward in my original paper (1991) has since been amply supported by several related developments.

The crucial idealization that plays a key role in classical physics is a “point.” Because of Heisenberg’s principle, $\Delta x \Delta p \geq \hbar/2$, quantum theory does not admit states with simultaneously vanishing Δx and Δp . However, as the study of the predictability sieve has demonstrated, in many situations relevant to the classical limit of quantum dynamics, one can expect decoherence to select pointer states that are localized in both Δx and Δp , that is, approximate minimum uncertainty wave packets. In effect, these wave packets are a quantum version of points, which appear naturally in the underdamped harmonic oscillator coupled weakly to the environment (Zurek et al. 1993, Gallis 1996). These results are also relevant to the transition from quantum to classical in the context of field theory with the added twist that the kinds of states selected will typically differ for bosonic and fermionic fields (Anglin and Zurek 1996) because bosons and fermions tend to couple differently to their environments. Finally, under suitable circumstances, einselection can even single out energy eigenstates of the self-Hamiltonian of the system, thus justifying in part the perception of “quantum jumps” (Paz and Zurek 1999).

An intriguing arena for the discussion of quantum-classical correspondence is quantum chaos. To begin with, classical and quantum evolutions from the same initial conditions of a system lead to very different phase-space “portraits.” The quantum phase-space portrait will depend on the particular representation used, but there are good reasons to favor the Wigner distribution. Studies that use the Wigner distribution indicate that, at the moment when quantum-classical correspondence is lost in chaotic dynamics, even the averages computed using properties of the classical and quantum states begin to differ (Karkuszewski et al. 2002).

Decoherence appears to be very effective in restoring correspondence. This point, originally demonstrated almost a decade ago (Zurek and Paz 1994, 1995) has since been amply corroborated by numerical evidence (Habib et al. 1998). Basically, decoherence eradicates the small-scale interference accompanying the rapid development of large-scale coherence in quantum ver-



sions of classically chaotic systems (refer to Figure A). This outcome was expected. In order for the quantum to classical correspondence to hold, the coherence length ℓ_C of the quantum state must satisfy the following inequality: $\ell_C = \hbar/(2D\lambda)^{1/2} \ll \chi$, where λ is the Lyapunov exponent, D is the usual coefficient describing the rate of decoherence, and χ is the scale on which the potential $V(x)$ is significantly nonlinear:

$$\chi \cong \sqrt{\frac{V'}{V'''}} .$$

When a quantum state is localized on scales small compared to χ (which is the import of the inequality above), its phase space evolution is effectively classical, but because of chaos and decoherence, it becomes irreversible and unpredictable. Nevertheless—as argued by Tanmoy Bhattacharya, Salman Habib, and Kurt Jacobs in the article “The Emergence of Classical Dynamics in a Quantum World” on page 110—one can even recover more or less classical trajectories by modeling a continuous measurement. However, this is an extra ingredient not in the spirit of the decoherence approach as it invokes the measurement process without explaining it.

A surprising corollary of this line of argument is the realization (Zurek and Paz 1994) that the dynamical second law—entropy production at the scale set by the

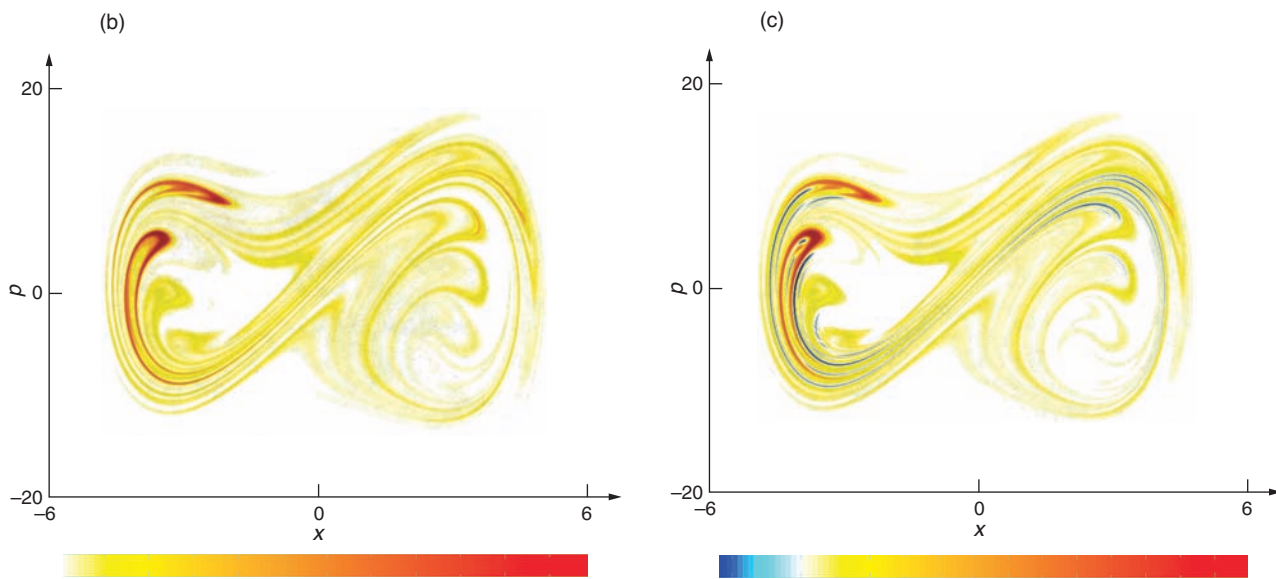


Figure A. Decoherence in a Chaotic Driven Double-Well System

This numerical study (Habib et al. 1998) of a chaotic driven double-well system described by the Hamiltonian $H = p^2/2m - Ax^2 + Bx^4 + Fx \cos(\omega t)$ with $m = 1$, $A = 10$, $B = 0.5$, $F = 10$, and $\omega = 6.07$ illustrates the effectiveness of decoherence in the transition from quantum to classical. These parameters result in a chaotic classical system with a Lyapunov exponent $\lambda \cong 0.5$. The three snapshots taken after 8 periods of the driving force illustrate phase space distributions in (a) the quantum case, (b) the classical case, and (c) the quantum case but with decoherence ($D = 0.025$). The initial condition was always the same Gaussian, and in the quantum cases, the state was pure. Interference fringes

are clearly visible in (a), which bears only a vague resemblance to the classical distribution in (b). By contrast, (c) shows that even modest decoherence helps restore the quantum-classical correspondence. In this example the coherence length ℓ_C is not much smaller than the typical nonlinearity scale, so the system is on the border between quantum and classical. Indeed, traces of quantum interference are still visible in (c) as blue “troughs,” or regions where the Wigner function is still slightly negative. The change in color from red to blue shown in the legends for (a) and (c) corresponds to a change from positive peaks to negative troughs. In the ab initio classical case (b), there are no negative troughs.

dynamics of the system and more or less independent of the strength of the coupling to the environment—is a natural and, indeed, an inevitable consequence of decoherence. This point has been since confirmed in numerical studies (Miller and Sarkar 1999, Pattanayak 1999, Monteoliva and Paz 2000).

Other surprising consequences of the study of Wigner functions in the quantum-chaotic context is the realization that they develop phase space structure on the scale associated with the sub-Planck action $a = \hbar^2/A \ll \hbar$, where A is the classical action of the system, and that this sub-Planck action is physically significant (Zurek 2001b). This can be seen in Figure A part (a), where a small black square with the area of \hbar is clearly larger than the smallest “ripples” in the image.

This point was to some extent anticipated by the plots of the Wigner functions of Schrödinger cats [see Figures 4(a) and 4(a') in this article] a version of which appeared in the 1991 *Physics Today* version of this paper—the interference term of the Wigner function has a sub-Planck structure.

A lot has happened in establishing phase-space aspects of quantum-classical correspondence, but a lot more remains to be done. (A more thorough summary of the past accomplishments and remaining goals can be found in Zurek 2001b).

Continued from page 101

Classical equations of motion are a necessary but insufficient ingredient of the classical limit: We must also obtain the correct structure of the classical phase space by barring all but the probability distributions of well-localized wave packets. The last term in Equation (24) has precisely this effect on nonclassical $W(x,p)$. For example, the Wigner function for the superposition of spatially localized wave packets—Figure 4(a)—has a sinusoidal modulation in the momentum coordinate produced by the oscillating term $\cos((\Delta x/\hbar)p)$. This term, however, is an eigenfunction of the diffusion operator $\partial^2/\partial p^2$ in the last term of Equation (24). As a result, the modulation is washed out by diffusion at a rate

$$\tau_D^{-1} = -\frac{\dot{W}}{W} = \frac{\left(D \frac{\partial^2}{\partial p^2} W \right)}{W} = \frac{2m\gamma k_B T (\Delta x)^2}{\hbar^2}. \quad (26)$$

Negative valleys of $W(x,p)$ fill in on a time scale of order τ_D , and the distribution retains just two peaks, which now correspond to two classical alternatives—see Figures 4(a) to 4(e). The Wigner function for a superposition of momenta, shown in Figure 4(a'), also decoheres as the dynamics causes the resulting difference in velocities to damp out the oscillations in position and again yield two classical alternatives—see Figures 4(b') to 4(e').

The ratio of the decoherence and relaxation time scales depends on \hbar^2/m —see Equation (19). Therefore, when m is large and \hbar small, τ_D can be nearly zero—decoherence can be nearly instantaneous—while, at the same time, the motion of small patches (which correspond to the probability distribution in classical phase space) in the smooth potential becomes reversible. This idealization is responsible for our confidence in classical mechanics, and, more generally, for many aspects of our belief in classical reality.

The discussion above demonstrates that decoherence and the transition from quantum to classical (usually regarded as esoteric) is an inevitable consequence of the immersion of a system in an environment. True, our considerations were based on a fairly specific model—a particle in a heat bath of harmonic oscillators. However, this is often a reasonable approximate model for many more complicated systems. Moreover, our key conclusions—such as the relation between the decoherence and relaxation time scales in Equation (19)—do not depend on any specific features of the model. Thus, one can hope that the viscosity and the resulting relaxation always imply decoherence and that the transition from quantum to classical can be always expected to take place on a time scale of the order of the above estimates.

Quantum Theory of Classical Reality

Classical reality can be defined purely in terms of classical states obeying classical laws. In the past few sections, we have seen how this reality emerges from the substrate of quantum physics: Open quantum systems are forced into states described by localized wave packets. They obey classical equations of motion, although with damping terms and fluctuations that have a quantum origin. What else is there to explain?

Controversies regarding the interpretation of quantum physics originate in the clash between the predictions of the Schrödinger equation and our perceptions. I will therefore conclude this paper by revisiting the source of the problem—our awareness of definite outcomes. If these mental processes were essentially unphysical, there would be no hope of formulating and addressing the ultimate question—why do we perceive just one of the quantum alternatives?—within the context of physics. Indeed, one might be tempted to follow Eugene Wigner (1961) and give consciousness the last word in collapsing the state vector. I shall assume the opposite. That is, I shall examine the idea that the higher mental processes all correspond to well-defined, but at present, poorly understood information-processing functions that are being carried out by physical systems, our brains.

Described in this manner, awareness becomes susceptible to physical analysis. In particular, the process of decoherence we have described above is bound to affect the states of the brain: Relevant observables of individual neurons, including chemical concentrations and electrical potentials, are macroscopic. They obey classical, dissipative equations of motion. Thus, any quantum superposition of the states of neurons will be destroyed far too quickly for us to become conscious of the quantum “goings on.” Decoherence, or more to the point, environment-induced superselection, applies to our own “state of mind.”

One might still ask why the preferred basis of neurons becomes correlated with the classical observables in the familiar universe. It would be, after all, so much easier to believe in quantum physics if we could train our senses to perceive nonclassical superpositions. One obvious reason is that the selection of the available interaction Hamiltonians is limited and constrains the choice of detectable observables. There is, however, another reason for this focus on the classical that must have played a decisive role: Our senses did not evolve for the purpose of verifying quantum mechanics. Rather, they have developed in the process in which survival of the fittest played a central role. There is no evolutionary reason for perception when nothing can be gained from prediction. And, as the predictability sieve illustrates, only quantum states that are robust in spite of decoherence, and hence, effectively classical, have predictable consequences. Indeed, classical reality can be regarded as nearly synonymous with predictability.

There is little doubt that the process of decoherence sketched in this paper is an important element of the big picture central to understanding the transition from quantum to classical. Decoherence destroys superpositions. The environment induces, in effect, a superselection rule that prevents certain superpositions from being observed. Only states that survive this process can become classical.

There is even less doubt that this rough outline will be further extended. Much work needs to be done both on technical issues (such as studying more realistic models that could lead to additional experiments) and on problems that require new conceptual input (such as defining what constitutes a “system” or answering the question of how an observer fits into the big picture).

Decoherence is of use within the framework of either of the two interpretations: It can supply a definition of the branches in Everett’s Many Worlds Interpretation, but it can also delineate the border that is so central to Bohr’s point of view. And if there is one lesson to be learned from what we already know about such matters, it is that information and its transfer play a key role in the quantum universe.

The natural sciences were built on a tacit assumption: Information about the universe can be acquired without changing its state. The ideal of “hard science” was to be objective and provide a description of reality. Information was regarded as unphysical, ethereal, a mere record of the tangible, material universe, an inconsequential reflection, existing beyond and essentially decoupled from the domain governed by the laws of physics. This view is no longer tenable (Landauer 1991). Quantum theory has put an end to this Laplacean dream about a mechanical universe. Observers of quantum phenomena can no longer be just passive spectators. Quantum laws make it impossible to gain information without changing the state of the measured object. The dividing line between what is and what is known to be has been blurred forever. While abolishing this boundary, quantum theory has simultaneously deprived the “conscious observer” of a monopoly on acquiring and storing information: Any correlation is a registration, any quantum state is a record of some other quantum state. When correlations are robust enough, or the record is sufficiently indelible, familiar classical “objective reality” emerges from the quantum substrate. Moreover, even a minute interaction with the environment, practically inevitable for any macroscopic object, will establish such a correlation: The environment will, in effect, measure the state of the object, and this suffices to destroy quantum coherence. The resulting decoherence plays, therefore, a vital role in facilitating the transition from quantum to classical.

The Existential Interpretation

The quantum theory of classical reality has developed significantly since 1991. These advances are now collectively known as the existential interpretation (Zurek 2001a). The basic difference between quantum and classical states is that the objective existence of the latter can be taken for granted. That is, a system's classical state can be simply "found out" by an observer originally ignorant of any of its characteristics. By contrast, quantum states are hopelessly "malleable"—it is impossible in principle for an observer to find out an unknown quantum state without perturbing it. The only exception to this rule occurs when an observer knows beforehand that the unknown state is one of the eigenstates of some definite observable. Then and only then can a nondemolition measurement (Caves et al. 1980) of that observable be devised such that another observer who knew the original state would not notice any perturbations when making a confirmatory measurement.

If the unknown state cannot be found out—as is indeed the case for isolated quantum systems—then one can make a persuasive case that such states are subjective, and that quantum state vectors are merely records of the observer's knowledge about the state of a fragment of the Universe (Fuchs and Peres 2000). However, einselection is capable of converting such malleable and "unreal" quantum states into solid elements of reality. Several ways to argue this point have been developed since the early discussions (Zurek 1993, 1998, 2001a). In effect, all of them rely on einselection, the emergence of the preferred set of pointer states. Thus, observers aware of the structure of the Hamiltonians (which are "objective," can be found out without "collateral damage", and in the real world, are known well enough in advance) can also divine the sets of preferred pointer states (if they exist) and thus discover the preexisting state of the system.

One way to understand this environment-induced objective existence is to recognize that observers—especially human observers—never measure anything directly. Instead, most of our data about the Universe is acquired when information about the systems of interest is intercepted and spread throughout the environment. The environment preferentially records the information about the pointer states, and hence, only information about the pointer states is readily available. This argument can be made more rigorous in simple models, whose redundancy can be more carefully quantified (Zurek 2000, 2001a).

This is an area of ongoing research. Acquisition of information about the systems from fragments of the environment leads to the so-called conditional quantum dynamics, a subject related to quantum trajectories (Carmichael 1993).

In particular one can show that the predictability sieve also works in this setting (Dalvit et al. 2001).

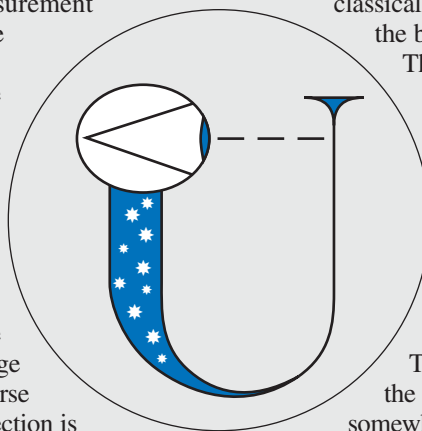
The overarching open question of the interpretation of quantum physics—the "meaning of the wave function"—appears to be in part answered by these recent developments. Two alternatives are usually listed as the only conceivable answers. The possibility that the state vector is purely epistemological (that is, solely a record of the observer's knowledge) is often associated with the Copenhagen Interpretation (Bohr 1928). The trouble with this view is that there is no unified description of the Universe as a whole: The classical domain of the Universe is a necessary prerequisite, so both classical and quantum theory are necessary and the border between them is, at best, ill-defined.

The alternative is to regard the state vector as an ontological entity—as a solid description of the state of the Universe akin to the classical states. But in this case (favored by the supporters of Everett's Many Worlds Interpretation), everything consistent with the universal state vector needs to be regarded as equally "real."

The view that seems to be emerging from the theory of decoherence is in some sense somewhere in between these two extremes.

Quantum state vectors can be real, but only when the superposition principle—a cornerstone of quantum behavior—is "turned off" by einselection. Yet einselection is caused by the transfer of information about selected observables. Hence, the ontological features of the state vectors—objective existence of the einselected states—is acquired through the epistemological "information transfer."

Obviously, more remains to be done. Equally obviously, however, decoherence and einselection are here to stay. They constrain the possible solutions after the quantum-classical transition in a manner suggestive of a still more radical view of the ultimate interpretation of quantum theory in which information seems destined to play a central role. Further speculative discussion of this point is beyond the scope of the present paper, but it will be certainly brought to the fore by (paradoxically) perhaps the most promising applications of quantum physics to information processing. Indeed, quantum computing inevitably poses questions that probe the very core of the distinction between quantum and classical. This development is an example of the unpredictability and serendipity of the process of scientific discovery: Questions originally asked for the most impractical of reasons—questions about the EPR paradox, the quantum-to-classical transition, the role of information, and the interpretation of the quantum state vector—have become relevant to practical applications such as quantum cryptography and quantum computation. ■



Acknowledgments

I would like to thank John Archibald Wheeler for many inspiring and enjoyable discussions on “the quantum” and Juan Pablo Paz for the pleasure of a long-standing collaboration on the subject.

Further Reading

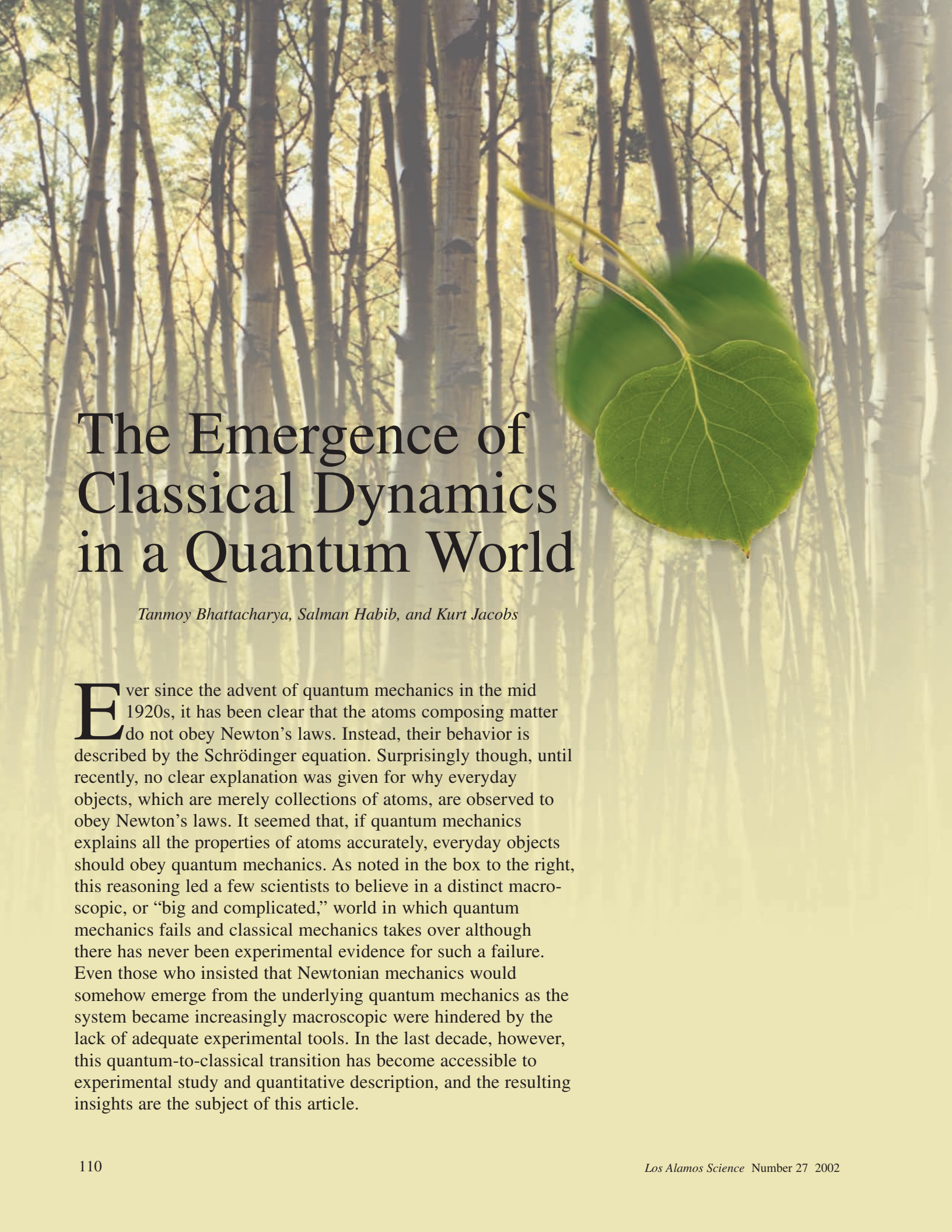
- Albrecht, A., 1992. Investigating Decoherence in a Simple System. *Phys. Rev. D* **46** (12): 5504.
- Anglin, J. R., and W. H. Zurek. 1996. Decoherence of Quantum Fields: Pointer States and Predictability. *Phys. Rev. D* **53** (12): 7327.
- Anglin, J. R., J. P. Paz, and W. H. Zurek. 1997. Deconstructing Decoherence. *Phys. Rev. A* **55** (6): 4041.
- Arndt, M., O. Nairz, J. VosAndreae, C. Keller, G. van der Zouw, A. Zeilinger. 1999. Wave-Particle Duality of C-60 Molecules. *Nature* **401** (6754): 680.
- Aspect, A., J. Dalibard, and G. Roger. 1982. Experimental Test of Bell’s Inequalities Using Time-Varying Analyzers. *Phys. Rev. Lett.* **49**: 1804.
- Aspect, A., P. Grangier, and G. Rogier. 1981. Experimental Tests of Realistic Local Theories via Bell’s Theorem. *Phys. Rev. Lett.* **47**: 460.
- Bell, J. S. 1964. On the Einstein Podolsky Rosen Paradox. *Physics* **1**: 195.
- Bohm, D. 1951. In *Quantum Theory*. Chap. 22, p. 611. Englewood Cliffs, NJ: Prentice Hall.
- Bohr, N. 1928. The Quantum Postulate and Recent Development of Atomic Theory. *Nature* **121**: 580.
Reprinted in *Quantum Theory and Measurement*. Edited by Wheeler, J. A., and W. H. Zurek. Princeton, NJ: Princeton University Press.
- Braginsky, V. B., Y. I. Vorontsov, and K. S. Thorne. 1980. Quantum Nondemolition Measurements. *Science* **209**: 547.
- Brune, M., E. Hagley, J. Dreyer, X. Maitre, C. Wunderlich, J. M. Raimond, and S. Haroche. 1996. Observing the Progressive Decoherence of the “Meter” in a Quantum Measurement. *Phys. Rev. Lett.* **77**: 4887.
- Caldeira, A. O., and A. J. Leggett. 1983a. Path Integral Approach to Quantum Brownian Motion. *Physica A*. **121**: 587.
- . 1983b. Quantum Tunneling in a Dissipative System. *Ann. Phys. (N. Y.)* **149** (2): 374.
- . 1985. Influence of Damping on Quantum Interference: An Exactly Soluble Model. *Phys. Rev. A* **31**: 1059.
- Carmichael, H. J. 1993. *An Open Systems Approach to Quantum Optics*. Berlin: Springer Verlag.
- Caves, C. M., K. S. Thorne, R. W. P. Dreuer, V. D. Sandberg, and M. Zimmerman. 1980. On the Measurement of a Weak Classical Force Coupled to a Quantum-Mechanical Oscillator. 1. Issues of Principle. *Rev. Mod. Phys.* **52**: 341.
- Chapman, M. S., T. D. Hammond, A. Lenef, J. Schmiedmayer, R. A. Rubenstein, E. Smith, and D. E. Pritchard. 1995. Photon Scattering from Atoms in an Atom Interferometer. *Phys. Rev. Lett.* **75** (21): 3783.
- Cheng, C. C., and M. G. Raymer. 1999. Long-Range Saturation of Spatial Decoherence in Wave-Field Transport in Random Multiple-Scattering Media. *Phys. Rev. Lett.* **82** (24): 4807.
- Dalvit, D. A. R., J. Dziarmaga, W. H. Zurek. 2001. Unconditional Pointer States from Conditional Master Equations. *Phys. Rev. Lett.* **86** (3): 373.
- Dekker, H. 1981. Classical and Quantum Mechanics of the Damped Harmonic Oscillator. *Phys. Rep.* **80**: 1.
- DeWitt, B. S. 1970. Quantum Mechanics and Reality. *Phys. Today* **23**: 30.
- DeWitt, B. S., and N. Graham, eds. 1973. *The Many-Worlds Interpretation of Quantum Mechanics*. Princeton: Princeton University Press.
- Einstein, A., B. Podolsky, and N. Rosen. 1935. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.* **47**: 777.
- Everett III, H. 1957. “Relative State” Formulation of Quantum Mechanics. *Rev. Mod. Phys.* **29**: 454.
- Feynman, R. P., and F. L. Vernon. 1963. The Theory of a General Quantum System Interacting with a Linear Dissipative System. *Ann. Phys.* **24**: 118.
- Friedman, J. R., V. Patel, W. Chen, S. K. Tolpygo, and J. E. Lukens. 2000. Quantum Superposition of Distinct Macroscopic States. *Nature* **406** (6791): 43.
- Fuchs, C. A., and A. Peres. 2000. Quantum Theory Needs No “Interpretation”. *Phys. Today* **53** (3): 70.
- Gallis, M. R. 1996. Emergence of Classicality via Decoherence Described by Lindblad Operators. *Phys. Rev. A* **53** (2): 655.
- Gell-Mann, M., and J. B. Hartle. 1990. Quantum Mechanics in the Light of Quantum Cosmology. In *Complexity, Entropy, and the Physics of Information*. p. 425. Edited by W. H. Zurek. Redwood City: Addison-Wesley.
- Griffiths, R. B. 1984. Consistent Histories and the Interpretation of Quantum Mechanics. *J. Stat. Phys.* **36**: 219.

- Haake, F., and D. F. Walls. 1986. In *Quantum Optics IV*. Edited by J. D. Harvey, and D. F. Walls. Berlin: Springer Verlag.
- Habib, S., K. Shizume, and W. H. Zurek. 1998. Decoherence, Chaos, and the Correspondence Principle. *Phys. Rev. Lett.* **80** (20): 4361.
- Haroche, S. 1998. Entanglement, Mesoscopic Superpositions and Decoherence Studies with Atoms and Photons in a Cavity. *Physica Scripta* **T76**: 159.
- Hartle, J. B. 1991. The Quantum Mechanics of Cosmology. In *Quantum Cosmology and Baby Universes: Proceedings of the 1989 Jerusalem Winter School*. Edited by S. Coleman, J. B. Hartle, T. Piran, and S. Weinberg. Singapore: World Scientific.
- Hu, B. L., J. P. Paz, and Y. Zhang. 1992. Quantum Brownian Motion in a General Environment: Exact Master Equation with Nonlocal Dissipation and Colored Noise. *Phys. Rev. D* **45**: 2843.
- Joos, E., and H. D. Zeh. 1985. The Emergence of Classical Properties Through Interaction with the Environment. *Z. Phys. B* **59**: 223.
- Karkuszewski, Z. P., J. Zakrzewski, and W. H. Zurek. 2002. Breakdown of Correspondence in Chaotic Systems: Ehrenfest Versus Localization Times. *Phys. Rev. A* **65** (4): 042113.
- Kokorowski, D. A., A. D. Cronin, T. D. Roberts, and D. E. Pritchard. 2001. From Single-to Multiple-Photon Decoherence in an Atom Interferometer. *Phys. Rev. Lett.* **86** (11): 2191.
- Landauer, R. 1991. Information is Physical. *Phys. Today* **44** (5): 23.
- Leggett, A. J., S. Chakravarty, A. T. Dorsey, M. P. A. Fisher, A. Garg, and W. Zwerger. 1987. Dynamics of the Dissipative System. *Rev. Mod. Phys.* **59**: 1.
- Milburn, G. J., and C. A. Holmes. 1986. Dissipative Quantum and Classical Liouville Mechanics of the Unharmonic Oscillator. *Phys. Rev. Lett.* **56**: 2237.
- Miller, P. A., and S. Sarkar. 1999. Signatures of Chaos in the Entanglement of Two Coupled Quantum Kicked Tops. *Phys. Rev. E* **60**: 1542.
- Monroe, C., D. M. Meekhof, B. E. King, and D. J. Wineland. 1996. A “Schrodinger Cat” Superposition State of an Atom. *Science* **272** (5265): 1131.
- Monteoliva, D., and J. P. Paz. 2000. Decoherence and the Rate of Entropy Production in Chaotic Quantum Systems. *Phys. Rev. Lett.* **85** (16): 3373.
- Mooij, J. E., T. P. Orlando, L. Levitov, L. Tian, C. H. van der Wal, and S. Lloyd. 1999. Josephson Persistent-Current Qubit. *Science* **285** (5430): 1036.
- Myatt, C. J., B. E. King, Q. A. Turchette, C. A. Sackett, D. Kielpinski, W. M. Itano, et al. 2000. Decoherence of Quantum Superpositions Through Coupling to Engineered Reservoirs. *Nature* **403**: 269.
- Ollivier, H., and W. H. Zurek. 2002. Quantum Discord: A Measure of the Quantumness of Correlations. *Phys. Rev. Lett.* **88** (1): 017901.
- Omnès, R. 1990. From Hilbert Space to Common Sense. *Ann. Phys.* **201**: 354.
- . 1992. Consistent Interpretation of Quantum Mechanics. *Rev. Mod. Phys.* **64**: 339.
- Pattanayak, A. K. 1999. Lyapunov Exponents Entropy Production and Decoherence. *Phys. Rev. Lett.* **83** (22): 4526.
- Paz, J. P., and W. H. Zurek. 1993. Environment-Induced Decoherence, Classicality, and Consistency of Quantum Histories. *Phys. Rev. D* **48** (6): 2728.
- . 1999. Quantum Limit of Decoherence: Environment Induced Superselection of Energy Eigenstates. *Phys. Rev. Lett.* **82** (26): 5181.
- . 2001. In *Coherent Atomic Matter Waves, Les Houches Lectures*. Edited by R. Kaiser, C. Westbrook, and F. Davids. Vol. 72, p. 533. Berlin: Springer.
- Paz, J. P., S. Habib, and W. H. Zurek. 1993. Reduction of the Wave Packet: Preferred Observable and Decoherence Time Scale. *Phys. Rev. D* **47**: 488.
- Pfau, T., S. Spälter, Ch. Kurtsiefer, C. R. Ekstrom, and J. Mlynek. 1994. Loss of Spatial Coherence by a Single Spontaneous Emission. *Phys. Rev. Lett.* **73** (9): 1223.
- Scully, M. O., B. G. Englert, and J. Schwinger. 1989. Spin Coherence and Humpty-Dumpty. III. The Effects of Observation. *Phys. Rev. A* **40**: 1775.
- Teich, M. C., and B. E. A. Saleh. 1990. Squeezed and Antibunched Light. *Phys. Today* **43** (6): 26.
- Tesche, C. D. 1986. Schroedinger’s Cat: A Realization in Superconducting Devices. *Ann. N. Y. Acad. Sci.* **480**: 36.
- Turchette, Q. A., C. J. Myatt, B. E. King, C. A. Sackett, D. Kielpinski, W. M. Itano, et al. 2000. Decoherence and Decay of Motional Quantum States of a Trapped Atom Coupled to Engineered Reservoirs. *Phys. Rev. A* **62**: 053807.

- Unruh, W. G., and W. H. Zurek. 1989. Reduction of a Wave Packet in Quantum Brownian Motion. *Phys. Rev. D* **40**:1071.
- Von Neumann, J. 1932. *Mathematische Grundlagen der Quanten Mechanik*. Berlin: Springer-Verlag. English translation by R. T. Beyer. 1955. *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press.
- Wheeler, J. A. 1957. Assessment of Everett's "Relative State" Formulation of Quantum Theory. *Rev. Mod. Phys.* **29**: 463.
- Wheeler, J. A., and W. H. Zurek, eds. 1983. *Quantum Theory and Measurement*. Princeton: Princeton University Press.
- Wigner, E. P. 1932. On the Quantum Correction for Thermodynamic Equilibrium. *Phys. Rev.* **40**: 749.
- . 1961. Remarks on the Mind-Body Question. In *The Scientist Speculates*. p. 284. Edited by I. J. Good. London: Heineman.
- . 1963. The Problem of Measurement. *Am. J. Phys.* **31**: 615.
- . 1983. In *Quantum Optics, Experimental Gravitation, and the Measurement Theory*. Edited by P. Meystre, and M. O. Scully. p. 43. New York: Plenum Press.
- Zeh, H. D. 1970. On the Interpretation of Measurement in Quantum Theory. *Found. Phys.* **1**: 69.
- Zurek, W. H. 1981. Pointer Basis of Quantum Apparatus: Into What Mixture Does the Wave Packet Collapse? *Phys. Rev. D* **24**: 1516.
- . 1982. Environment-Induced Superselection Rules. *Phys. Rev. D* **26**: 1862.
- . 1984. Reduction of the Wave Packet: How Long Does It Take? In *Frontiers of Nonequilibrium Statistical Physics*. Edited by P. Meystre, and M. O. Scully. New York: Plenum.
- . 1991. Decoherence and the Transition From Quantum to Classical. *Phys. Today* **44** (10): 36.
- . 1993. Preferred States, Predictability, Classicality, and the Environment-Induced Decoherence. *Prog. Theor. Phys.* **89** (2): 281.
- . 1998. Decoherence, Chaos, Quantum-Classical Correspondence, and the Algorithmic Arrow of Time. *Physica Scripta* **T76**: 186.
- . 2000. Einselection and Decoherence from an Information Theory Perspective. *Ann. Phys. (Leipzig)* **9** (11–12): 855.
- . 2001a. Decoherence, Einselection, and the Quantum Origins of the Classical. [Online]: [http://eprints.lanl.gov \(quant-ph/0105127\)](http://eprints.lanl.gov (quant-ph/0105127)).
- . 2001b. Sub-Planck Structure in Phase Space and its Relevance for Quantum Decoherence. *Nature* **412**: 712.
- Zurek, W. H., and J. P. Paz. 1994. Decoherence, Chaos, and the Second Law. *Phys. Rev. Lett.* **72** (16): 2508.
- . 1995. Quantum Chaos: A Decoherent Definition. *Physica D* **83** (1–3): 300.
- Zurek, W. H., S. Habib, and J. P. Paz. 1993. Coherent States via Decoherence. *Phys. Rev. Lett.* **70** (9): 1187.



Wojciech Hubert Zurek was educated in Kraków, in his native Poland (M. Sc., 1974) and in Austin, Texas (Ph. D. in physics, 1979). He was a Richard Chace Tolman fellow at the California Institute of Technology and a J. Robert Oppenheimer postdoctoral fellow at the Los Alamos National Laboratory, where he became a technical staff member. Between 1990 and 1996, Wojciech led the Theoretical Astrophysics Group. In 1996, he was selected as a Los Alamos National Laboratory Fellow. He is a Foreign Associate of the Cosmology Program of the Canadian Institute of Advanced Research and the founder of the Complexity, Entropy, and Physics of Information Network of the Santa Fe Institute. His research interests include decoherence, physics of quantum and classical information, foundations of statistical and quantum physics and astrophysics.



The Emergence of Classical Dynamics in a Quantum World

Tanmoy Bhattacharya, Salman Habib, and Kurt Jacobs

Ever since the advent of quantum mechanics in the mid 1920s, it has been clear that the atoms composing matter do not obey Newton's laws. Instead, their behavior is described by the Schrödinger equation. Surprisingly though, until recently, no clear explanation was given for why everyday objects, which are merely collections of atoms, are observed to obey Newton's laws. It seemed that, if quantum mechanics explains all the properties of atoms accurately, everyday objects should obey quantum mechanics. As noted in the box to the right, this reasoning led a few scientists to believe in a distinct macroscopic, or "big and complicated," world in which quantum mechanics fails and classical mechanics takes over although there has never been experimental evidence for such a failure. Even those who insisted that Newtonian mechanics would somehow emerge from the underlying quantum mechanics as the system became increasingly macroscopic were hindered by the lack of adequate experimental tools. In the last decade, however, this quantum-to-classical transition has become accessible to experimental study and quantitative description, and the resulting insights are the subject of this article.

A Historical Perspective

The demands imposed by quantum mechanics on the disciplines of epistemology and ontology have occupied the greatest minds. Unlike the theory of relativity, the other great idea that shaped physical notions at the same time, quantum mechanics does far more than modify Newton's equations of motion. Whereas relativity redefines the concepts of space and time in terms of the observer, quantum mechanics denies an aspect of reality to system properties (such as position and momentum) until they are measured. This apparent creation of reality upon measurement is so profound a change that it has engendered an uneasiness defying formal statement, not to mention a solution. The difficulties are often referred to as "the measurement problem." Carried to its logical extreme, the problem is that, if quantum mechanics were the ultimate theory, it could deny any reality to the measurement results themselves unless they were observed by yet another system, ad infinitum. Even the pioneers of quantum mechanics had great difficulty conceiving of it as a fundamental theory without relying on the existence of a classical world in which it is embedded (Landau and Lifshitz 1965).

Quantum mechanics challenges us on another front as well. From our intuitive understanding of Bayes' theorem for conditional probability, we constantly infer the behavior of systems that are observed incompletely. Quantum mechanics, although probabilistic, violates Bayes' theorem and thereby our intuition. Yet the very basis for our concepts of space and time and for our intuitive Bayesian view comes from observing the natural world. How come the world appears to be so classical when the fundamental theory describing it is manifestly not so? This is the problem of the quantum-to-classical transition treated in this article.

One of the reasons the quantum-to-classical transition took so long to come under serious investigation may be that it was confused with the measurement problem. In fact, the problem of assigning intrinsic reality to properties of individual quantum systems gave rise to a purely statistical interpretation of quantum mechanics. In this view, quantum laws apply only to ensembles of identically prepared systems.

The quantum-to-classical transition may also have been ignored in the early days because regular, rather than chaotic, systems were the subject of interest. In the former systems, individual trajectories carry little information, and quantization is straightforward. Even though Henri Poincaré (1992) had understood the key aspects of chaos and Albert Einstein (1917) had realized its consequences for the Bohr-Sommerfeld quantization schemes, which were popular at that time, this subject was never in the spotlight, and interest in it was not sustained until fairly recently.

As experimental technology progressed to the point at which single quanta could be measured with precision,

the façade of ensemble statistics could no longer hide the reality of the counterclassical nature of quantum mechanics. In particular, a vast array of quantum features, such as interference, came to be seen as everyday occurrences in these experiments.

Many interpretations of quantum mechanics developed. Some appealed to an anthropic principle, according to which life evolved to interpret the world classically, others imagined a manifold of universes, and yet others looked for a set of histories that were consistent enough for classical reasoning to proceed (Omnès 1994, Zurek in this issue). However, by themselves, these approaches do not offer a dynamical explanation for the suppression of interference in the classical world. The key realization that led to a partial understanding of the classical limit was that weak interactions of a system with its environment are universal (Landau and Lifshitz 1980) and remove the nonclassical terms in the quantum evolution (Zurek 1991). The folklore developed that this was the the only effect of a sufficiently weak interaction in almost any system. In fact, Wigner functions (the closest quantum analogues to classical probability distributions in phase space) did often become positive, but they failed to become localized along individual classical trajectories. In the heyday of ensemble interpretations, this was not a problem because classical ensembles would have been represented by exactly such distributions. When applied to a single quantum system in a single experiment, however, this delocalized positive distribution is distinctly dissatisfying.

Furthermore, even when a state is describable by a positive distribution, it is not obvious that the dynamics can be interpreted as the dynamics of any classical ensemble without hypothesizing a multitude of "hidden" variables (Schack and Caves 1999). And finally, the original hope that a weak interaction merely erases interference turned out to be untenable, at least in some systems (Habib et al. 2000).

The underlying reason for environmental action to produce a delocalized probability distribution is that even if we take a single classical system with its initial (or subsequent) positions unknown, our state of knowledge can be encoded by that distribution. But in an actual experiment, we do know the position of the system because we continuously measure it. Without this continuous (or almost continuous) measurement, we would not have the concept of a classical trajectory. And without a classical trajectory, such remarkable signals of chaos as the Lyapunov exponent would be experimentally immeasurable. These developments brought us to our current view that continuous measurements provide the key to understanding the quantum-to-classical transition.

We will illustrate the problems involved in describing the quantum-to-classical transition by using the example of a baseball moving through the air. Most often, we describe how the ball moves through air, how it spins, or how it deforms. Regardless of which degree of freedom we might consider—whether it is the position of the center of mass, angular orientation, or deviation from sphericity—in the final analysis, those variables are merely a combination of the positions (or other properties) of the individual atoms. As all the properties of each of these atoms, including position, are described by quantum mechanics, how is it that the ball as a whole obeys Newton's equation instead of some averaged form of the Schrödinger equation?

Even more difficult to explain is how the chaotic behavior of classical, nonlinear systems emerges from the behavior of quantum systems. Classical, nonlinear, dynamical systems exhibit extreme sensitivity to initial conditions. This means that, if the initial states of two identical copies of a system (for example, particle positions and momenta) differ by some tiny amount, those differences magnify with time at an exponential rate. As a result, in a very short time, the two systems follow very different evolutionary paths. On the other hand, concepts such as precise position and momentum do not make sense according to quantum mechanics: We can describe the state of a system in terms of these variables only probabilistically. The Schrödinger equation governing the evolution of these probabilities typically makes the probability distributions diffuse over time. The final state of such systems is typically not very sensitive to the initial conditions, and the systems do not exhibit chaos in the classical sense.

The key to resolving these contradictions hinges on the following observation: While macroscopic mechanical systems may be described by single quantum degrees of freedom, those variables are subject to observation and interaction with their environment, which are continual influences. For example, a baseball's center-of-mass coordinate is continually affected by the numerous properties of the atoms composing the baseball, including thermal motion, the air that surrounds it, which is also in thermal motion, and the light that reflects off it. The process of observing the baseball's motion also involves interaction with the environment: Light reflected off the baseball and captured by the observer's eye creates a trace of the motion on the retina.

In the next section, we will show that, under conditions that refine the intuitive concept of what is macroscopic, the motion of a quantum system is basically indistinguishable from that of a classical system! In effect, observing a quantum system provides information about it and counteracts the inherent tendency of the probability distribution to diffuse over time although observation creates an irreducible disturbance. In other words, as we see the system continuously, we know where it is and do not have to rely upon the progressively imprecise theoretical predictions of where it could be. When one takes into account this "localization" of the probability distribution encoding our knowledge of the system, the equations governing the expected measurement results (that is, the equations telling us what we observe) become nonlinear in precisely the right way to recover an approximate form of classical dynamics—for example, Newton's laws in the baseball example.

What happens when no one observes the system? Does the baseball suddenly start behaving quantum mechanically if all observers close their eyes? The answer is hidden in a simple fact: Any interaction with a sufficiently complicated external world has the same effect as a series of measurements whose results are not recorded. In other words, the nature of the disturbance on the system due to the system's interactions with the external world is identical to that of the disturbance observed as an irreducible component of measurement. Naturally, questions about the path of the baseball can't be verified if there are no observers, but other aspects of its classical nature can, and do, survive.

Classical vs Quantum Trajectories

Let us now turn to some significant details. To describe the motion of a single classical particle, all we need to do is specify a spatially dependent, and possibly time-dependent, force that acts on the particle and substitute it into Newton's equations. The resulting set of two coupled differential equations, one for the position x of the particle and the other for the momentum p , predicts the evolution of the particle's state. If the force on the particle is denoted by $F(x,t)$, the equations of motion are

$$\dot{x} = \frac{p}{m} , \quad (1)$$

and

$$\dot{p} = F(x,t) = -\partial_x V(x,t) , \quad (2)$$

where V is the potential.

To visualize the motion, one can plot the particle's position and momentum as they change in time. The resulting curve is called a trajectory in phase space (see Figure 1). The axes of phase space delineate the possible spatial and momentum coordinates that the single particle can take. A classical particle's state is given at any time by a point in phase space, and its motion therefore traces out a curve, or trajectory, in phase space.

By contrast, the state of a quantum particle is not described by a single point in phase space. Because of the Heisenberg uncertainty principle, the position and momentum cannot simultaneously be known with arbitrary precision, and the state of the system must therefore be described by a kind of probability density in phase space. This pseudoprobability function is called the Wigner function and is denoted by $f_W(x,p)$. As expected for a true probability density, the integral of the Wigner function over position gives the probability density for p , and the integral over p gives the probability density for x . However, because the Wigner function may be negative in places, we should not try to interpret it too literally. Be that as it may, when we specify the force on the particle, $F(x,t)$, the evolution of the Wigner function is given by the quantum Liouville equation, which is

$$\begin{aligned} \dot{f}_W(x,p) = & - \left[\frac{p}{m} \partial_x + F(x,t) \partial_p \right] f_W(x,p) \\ & + \sum_{\lambda=1}^{\infty} \frac{1}{(2\lambda+1)!} \left(\frac{\hbar}{2i} \right)^{2\lambda} \partial_x^{2\lambda+1} V(x,t) \partial_p^{2\lambda+1} f_W(x,p) . \end{aligned} \quad (3)$$

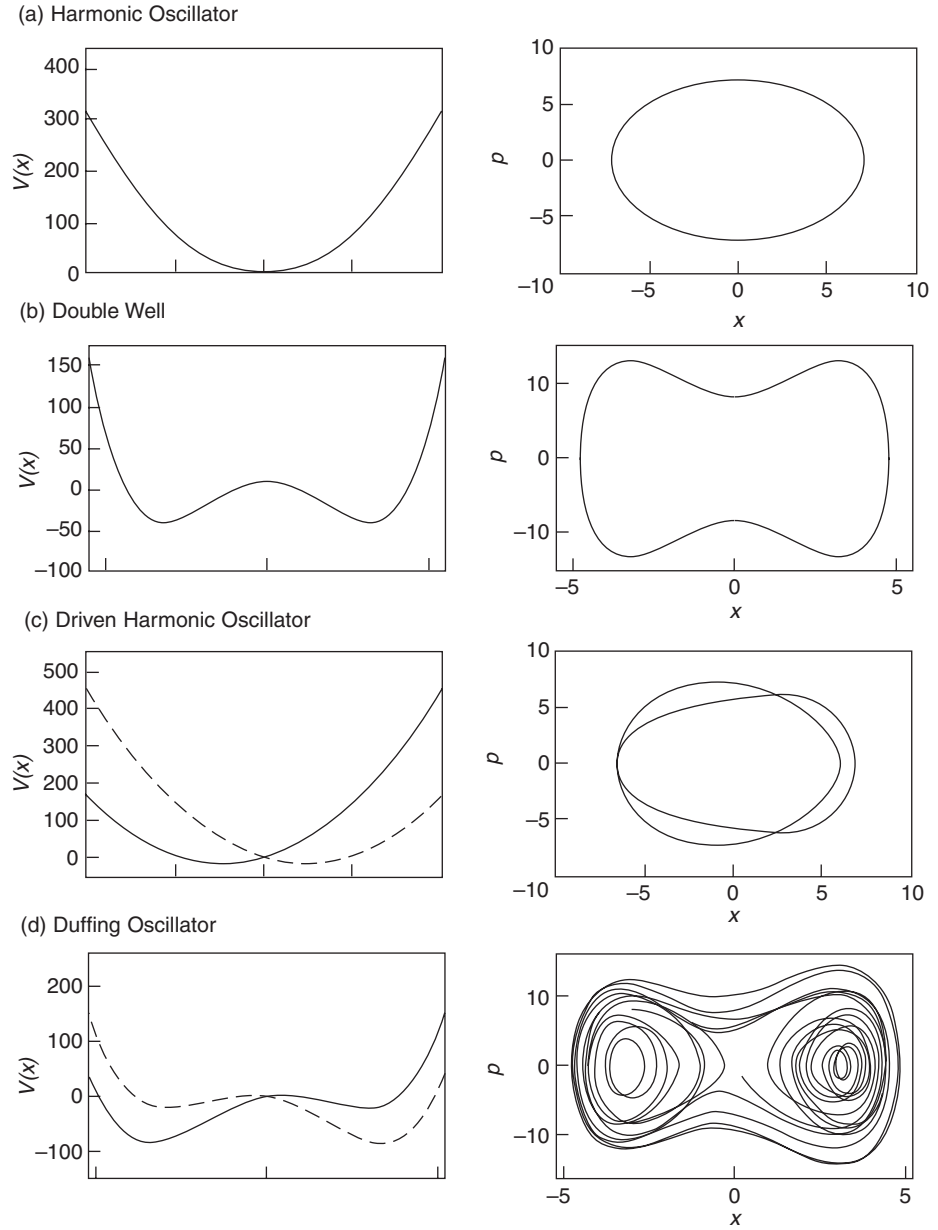
Clearly, in order for a quantum particle to behave as a classical particle, we must be able to assign it a position and momentum, even if only approximately. For example, if the Wigner function stays localized in phase space throughout its evolution, then the centroid of the Wigner function¹ could be interpreted at each time as the location of the particle in phase space.

¹ The centroid of the Wigner function is the point in phase space consisting of the mean values of x and p , that is $(\langle x \rangle, \langle p \rangle)$.

Figure 1. Potentials and Phase-Space Trajectories for Single-Particle Systems

The figure shows four systems in which a single particle is constrained to move in a one-dimensional potential. The four systems are (a) a harmonic oscillator, (b) a double well, (c) a driven harmonic oscillator, and (d) a driven double well, also known as a Duffing oscillator.

As the potentials increase in complexity from (a) to (d), so do the phase-space trajectories. In (c) and (d), the potential is time dependent, oscillating back and forth between the solid and dashed curves during each period. In (d), the force is nonlinear, and the trajectory covers increasingly more of the phase space as time passes.



Moreover, the Liouville equation yields the following equations of motion for the centroid:

$$\langle \dot{x} \rangle = \frac{\langle p \rangle}{m} , \tag{4}$$

and

$$\langle \dot{p} \rangle = \langle F(x,t) \rangle , \tag{5}$$

where m is the mass of the particle. This result, referred to as Ehrenfest's theorem,² says that the equations of motion for the centroid formally resemble the classical ones but differ from classical dynamics in that the force F has been replaced with the average value of F over the Wigner function. Suppose again that the Wigner function is sharply peaked about $\langle x \rangle$ and $\langle p \rangle$. In that case, we can approximate $\langle F(x) \rangle$ as a Taylor series expansion about $\langle x \rangle$:

$$\langle F(x) \rangle = F(\langle x \rangle) + \frac{\sigma_x^2}{2} \partial_x^2 F(\langle x \rangle) + \dots, \quad (6)$$

where σ_x^2 is the variance of x so that $\sigma_x^2 = \langle (x - \langle x \rangle)^2 \rangle$. If the second and higher terms in the Taylor expansion are negligible, the equations for the centroid become

$$\langle \dot{x} \rangle = \frac{\langle p \rangle}{m}, \quad (7)$$

and

$$\langle \dot{p} \rangle = F(\langle x \rangle, t). \quad (8)$$

And these equations for the centroid are identical to the equation of motion for the classical particle! If we somehow arrange to start the system with a sharply localized Wigner function, the motion of the centroid will start out by being classical, and Equation (6) indicates precisely how sharply peaked the Wigner function needs to be.

However, the Wigner function of an unobserved quantum particle rarely remains localized even if for some reason it starts off that way. In fact, when an otherwise noninteracting quantum particle is subject to a nonlinear force, that is, a force with a nonlinear dependence on x , the evolution usually causes the Wigner function to develop a complex structure and spread out over large areas of phase space. In the sequence of plots in Figure 2(a-d), the Wigner function is shown to spread out in phase space under the influence of a nonlinear force. Once the Wigner function has spread out in this way, the evolution of the centroid bears no resemblance to a classical trajectory.

So, the key issue in understanding the quantum-to-classical transition is the following: Why should the Wigner function localize and stay localized thereafter? As stated in the introduction, this is an outcome of continuous observation (measurement). We therefore now turn to the theory of continuous measurements.

Continuous Measurement

In simple terms, any process that yields a continuous stream of information may be termed continuous observation. Because in quantum mechanics measurement creates an irreducible disturbance on the observed system and we do not wish to disturb the system unduly, the desired measurement process must yield a limited amount of information in a finite time. Simple projective measurements, also known as von Neumann

² According to Ehrenfest's theorem, a quantum-mechanical wave packet obeys the equation of motion of the corresponding classical particle when the position, momentum, and force acting on the particle are replaced by the expectation values of these quantities.

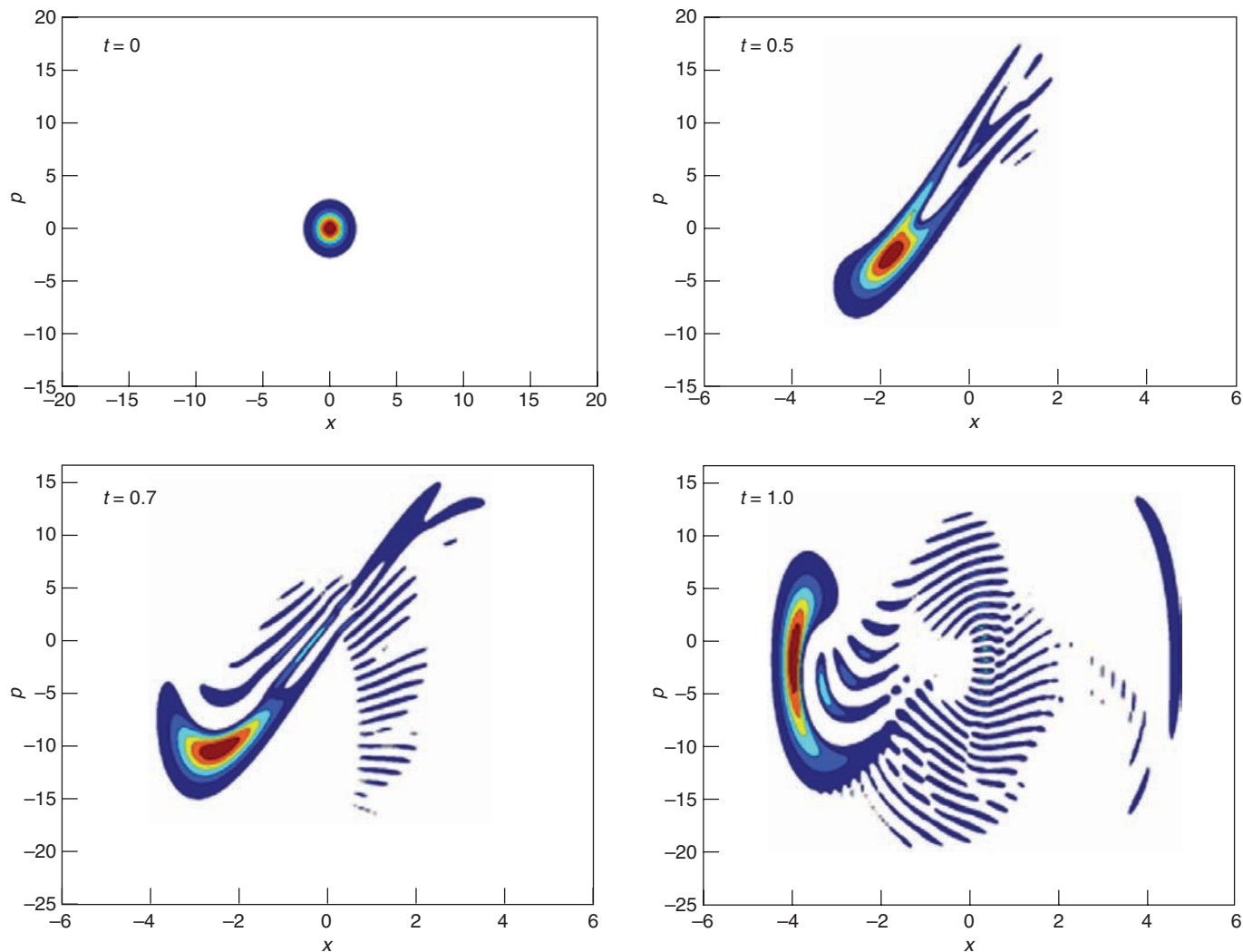


Figure 2. Evolution of the Wigner Function under a Nonlinear Force

These four snapshots show the Wigner function at different times during the simulations of the Duffing oscillator. At $t = 0$, the Wigner function is localized around a single point. As time passes, however, the Wigner function becomes increasingly delocalized under the nonlinear potential of the Duffing oscillator.

measurements, introduced in undergraduate quantum mechanics courses, are not adequate for describing continuous measurements because they yield complete information instantaneously. The proper description of measurements that extract information continuously, however, results from a straightforward generalization of von Neumann measurements (Davies 1976, Kraus 1983, Carmichael 1993). All we need to do is let the system interact weakly with another one, such as a light beam, so that the state of the auxiliary system should gather very little information about the main one over short periods and thereby the system of interest should be perturbed only slightly. Only a very small part of the information gathered by a projective measurement of the auxiliary system then pertains to the system of interest, and a continuous limit of this measurement process can then be taken. By the mid 1990s, this generalization of the standard measurement theory was already being used to describe continuous position measurement by laser beams. In our analysis, we use the methods developed as part of this effort.

A simple, yet sufficiently realistic, analogy to measuring position by direct observation is measuring the position of a moving mirror by reflecting a laser beam off the mirror and continuously monitoring the phase of the reflected light. As the knowledge of the system is initially imprecise, there is a random component in the measurement record. Classically, our knowledge of the system state may be refined to an arbitrary accuracy over time, and the random component is thereby reduced. Quantum mechanically,

however, the measurement itself disturbs the system, and our knowledge cannot be improved arbitrarily. As a result, the measurement record continues to have a random component.

An equivalent way of understanding this random component is to note that the measurement process may be characterized by the rate at which information is obtained. A more powerful measurement is one in which information is obtained at a faster rate. Because of the Heisenberg uncertainty relation, if we obtain information about position, we lose information about momentum. But uncertainty in momentum turns into uncertainty in position at the very next instant. This random feedback guarantees that a continuous measurement will cause the system to be driven by noise: The higher the rate at which information is obtained, the more the noise. For a position measurement, the rate of information extraction is usually characterized by a constant, k , that measures how fast the precision in our knowledge of position, $1/\sigma_x^2$, would increase per unit time in the absence of the accompanying disturbance. In the laser measurement of position, k is determined by the power of the laser. The more powerful the laser, the stronger the measurement, and the more noise introduced by the photon collisions.

Now we are in a position to see how and under what circumstances continuous measurement transforms quantum into classical dynamics, resulting in the quantum-to-classical transition. We can include the effects of the observation on the motion of the particle by writing down a stochastic Liouville equation, that is, a Liouville equation with a random component. This equation is given in the box “Conditions for Approximate Classical Motion under Continuous Measurement” on the next page. The resulting equations of motion for the centroid of the Wigner function are

$$\langle \dot{x} \rangle = \frac{\langle p \rangle}{m} + \sqrt{8k}\sigma_x^2 \xi(t) , \tag{9}$$

and

$$\langle \dot{p} \rangle = F(\langle x \rangle, t) + \sqrt{8k}C_{xp}\xi(t) , \tag{10}$$

where $C_{xp} = (1/2)(\langle xp \rangle + \langle px \rangle - 2\langle x \rangle \langle p \rangle)$ is the covariance of x and p , and $\xi(t)$ is a Gaussian white noise.³

We have now reached the crux of the quantum-to-classical transition. To keep the Wigner function well localized, a strong measurement, or a large k , is needed. But Equations (9) and (10) show that a strong measurement introduces a lot of noise. In classical mechanics, however, we deal with systems in which the amount of noise, if any, is imperceptible compared with the scale of the distances traveled by the particle. We must therefore determine the circumstances under which continuous measurement will maintain sufficient localization for the classical equations to be approximately valid without introducing a level of noise that would affect this scale of everyday physics.

³ White noise is random noise that has constant energy per unit bandwidth at every frequency. In reality, the actual recording of the measurement always occurs at a finite rate. So, effectively, the white noise gets filtered through a low-pass filter, which cuts out high frequencies.

Conditions for Approximate Classical Motion

The evolution of the Wigner function f_W for a single particle subjected to a continuous measurement of position is given by the stochastic Liouville equation:

$$f_W(x, p, t + dt) = \left[1 - dt \left[\left(\frac{p}{m} \right) \partial_x + F(x, t) \partial_p \right] + dt \sum_{\lambda=1}^{\infty} \frac{1}{(2\lambda+1)!} \left(\frac{\hbar}{2i} \right)^{2\lambda} \partial_x^{2\lambda+1} V(x, t) \partial_p^{2\lambda+1} \right] f_W(x, p) + \sqrt{8k} dt \xi (x - \langle x \rangle) f_W(x, p) , \quad (1)$$

where F is the force on the particle, ξ is a Gaussian white noise, and k is a constant characterizing the rate of information extraction. Making a Gaussian approximation for the Wigner function, which according to numerical studies is a good approximation when localization is maintained by the measurement, the equations of motion for the variances of x and p , σ_x^2 and σ_p^2 , are

$$\begin{aligned} \dot{\sigma}_x^2 &= \frac{2}{m} C_{xp} - 8k \sigma_x^4 , \quad \text{and} \\ \dot{\sigma}_p^2 &= 2\hbar^2 k - 8k C_{xp}^2 + 2\partial_x F C_{xp} , \quad \text{where the covariance of } x \text{ and } p \text{ is } C_{xp} = \frac{1}{2} (\langle xp \rangle + \langle px \rangle - 2\langle x \rangle \langle p \rangle) . \\ \dot{C}_{xp} &= \frac{1}{m} \sigma_p^2 - 8k \sigma_x^2 C_{xp} + \partial_x F \sigma_x^2 . \end{aligned} \quad (2)$$

the noise has negligible effect in these equations when the Wigner function stays Gaussian.

First, we solve these equations for the steady state and then impose on this solution the conditions required for classical dynamics to result. In order for the Wigner function to remain sufficiently localized, the measurement strength k must stop the spread of the wave function at the unstable points, $\partial_x F > 0$.*

$$8k \gg \left| \frac{\partial_x^2 F}{F} \right| \sqrt{\frac{|\partial_x F|}{2m}} . \quad (3)$$

If noise is to bring about only a negligible perturbation to the classical dynamics, it is sufficient that, at a typical point on the trajectory, the measurement satisfy

$$\frac{2|\partial_x F|}{s} \ll \hbar k \ll \frac{|\partial_x F|s}{4} , \quad (4)$$

where s is the typical value of the system's action[†] in units of \hbar . Obviously, as s becomes much larger than $2\sqrt{2}$ this relationship is satisfied for an ever-larger range of k . At the spot where this range is sufficiently large, we obtain the classical limit.

* If the nonlinearity is large on the quantum scale, $\hbar \left| \frac{\partial_x^2 F}{F} \right| \geq 4\sqrt{m|\partial_x F|}$, then $8k$ needs to be much larger than $(\partial_x^2 F)^2 \hbar / 4mF^2$ irrespective of the sign of $\partial_x F$. This observation does not change the argument in the body of the paper.

† We are assuming that both $[mF^2/(\partial_x F)^2]|F/p|$ and $E|p/4F|$ evaluated at a typical point of the trajectory are comparable to the action of the system, and we define that action to be $\hbar s$.

With analytical tools alone, this problem cannot be solved. However, one can take a semianalytical approach by accepting two important results that come from numerical simulations: (1) Any Wigner function localizes under a sufficiently strong measurement, and (2) under such a measurement, once the Wigner function becomes localized, it is approximately described by a narrow Gaussian at all later times. Therefore, we assume a Gaussian form for the Wigner function, write the equations determining how the variances and covariances change with time, and solve those equations to find their values in a steady state. Having all these ingredients, we can then find the conditions under which the noise terms are small and the system remains well localized (see the box on the opposite page). Our central conclusion is that a quantum system will behave almost classically for an ever-increasing range of measurement strengths when the action of the system is large compared with the reduced Planck constant \hbar .

This concept may be understood heuristically in the following way: Because of the uncertainty principle, the effective area where the localized Gaussian Wigner function is nonzero can never be less than \hbar . If this limiting area is so large compared with the scale of the problem that it cannot be considered localized, we certainly do not expect classical behavior. Conversely, as long as the measurement extracts information at a sufficiently low rate to avoid squeezing the Wigner function to a smaller scale than the limiting one, the quantum noise remains on the scale of the variances themselves. As a result, the system behaves almost classically.

There are systems, however, whose phase space is sufficiently small for quantum effects to be manifest or even dominant. This is true, for example, of isolated spin systems with small total angular momenta. Even when they are observed and interacting with the environment, these spin systems are expected to be indescribable by the classical laws of motion. A spin coupled to other degrees of freedom such as position is a more interesting case, especially when the position of the system would have followed a classical trajectory in the absence of that interaction. To what extent, if at all, that coupling stops position from following a classical trajectory is the subject of ongoing research (Ghose et al. 2002).

Chaos in a Quantum System under Continuous Observation

As an illustration of these general ideas, we consider the Duffing oscillator, a single particle sitting in a double-well potential and driven sinusoidally—see Figure 1(d). We chose this nonlinear system because it has been studied in depth and it allows us to choose parameters that produce chaotic behavior over most of the system's phase space. Our test will indicate whether chaotic classical motion is a good approximate description of this quantum system when it is under continuous observation. To diagnose the presence of chaos, we calculate the maximal Lyapunov exponent, the most rigorous measure of chaotic behavior,⁴ and compare our calculated value for the quantum system with the classical value.

The Hamiltonian for the particle in the double-well potential is

$$H = \frac{p^2}{2m} + Bx^4 - Ax^2 + \Lambda x \cos(\omega t), \quad (11)$$

where m , A , B , Λ , and ω are parameters that determine the size of the particle and the

⁴ The maximal Lyapunov exponent is one of a number of coefficients that describe the rates at which nearby trajectories in phase space converge or diverge.

spatial extent of the phase space. The action should be large enough so that the particle can behave almost classically, yet small enough to illustrate how tiny it needs to be before quantum effects on the particle become dominant. Bearing this requirement in mind, we choose a mass $m = 1$ picogram, a spring constant $A = 0.99$ piconewton per meter, a nonlinearity $A/B = 0.02$ square micrometer, a peak driving force of $\lambda = 0.03$ attonewton, and a driving frequency $\omega = 60$ rad per second. Because of the weakness of the nonlinearity, the distance between the two minima of the double well is only about 206 nanometers, and the height of the potential is only 33 nano-electronvolts. The frequency of the driving force is 10 hertz. For these values, a measurement strength k of 93 per square picometer per second, which corresponds to a laser power of about 0.24 microwatt, is adequate to keep the motion classical, or the Wigner function well localized.

To study the system numerically, we allow the particle's Wigner function to evolve according to the stochastic Liouville equation for approximately 50 periods of the driving force and then check that it remains well localized in the potential. We find, indeed, that the width of the Wigner function in position (given by the square root of the position variance σ_x^2) is always less than 2 nanometers. Thus the position of the particle is always well resolved by the measurement as the system evolves. In addition, an inspection of the centroid's trajectory shows that the noise is negligible. In order to verify that the motion is, in fact, that of a classical Duffing oscillator, we perform two tests. The first is to plot a stroboscopic map showing the particle's motion in phase space and then compare that map with the corresponding one of the classical Duffing oscillator driven by a small amount of noise. The observed quantum map and the classical map are displayed in Figure 3.

The two stroboscopic maps are very similar and show qualitatively that the quantum dynamics under continuous measurement exhibits chaotic behavior analogous to classical chaos. To verify this finding quantitatively, we conduct a second test and calculate the Lyapunov exponent for both systems. As we already mentioned, trajectories that are initially separated by a very small phase-space distance, $d(0)$, diverge exponentially as a function of time in chaotic systems. The Lyapunov exponent λ , which determines the rate of this exponential increase, is defined to be

$$\lambda = \lim_{t \rightarrow \infty} \lim_{d(0) \rightarrow 0} \frac{\ln d(t)}{t} . \quad (12)$$

To calculate this exponent, we first choose a single fiducial trajectory in which the centroid of the Wigner function starts at the phase-space point given by $\langle x \rangle = -98$ nanometers and $\langle p \rangle = 2.6$ picograms micrometers per second (pg $\mu\text{m/s}$). At 17 intervals along this trajectory, each separated by approximately 20 periods of the driving force, we obtain neighboring trajectories by varying the noise realization. We calculate how these trajectories diverge from the initial trajectory and average the difference over the 17 sample trajectories. We then carry out this procedure for 10 fiducial trajectories, all starting at the same initial point but differing because of different noise realizations. Plotting the logarithm of this average divergence as a function of time results in a line whose slope is the Lyapunov exponent. In Figure 4, we plot the logarithm of the average divergence for both the observed quantum system and the classical system driven with a small amount of noise. The slope of the lines drawn through the curves gives the Lyapunov exponent, which in both cases is 5.7(2) per second. To show that the noise has a negligible effect on the dynamics, we also calculate the Lyapunov exponent for the classical system with no noise, using trajectories starting in a small region around the point given by $x = -98$ nanometers

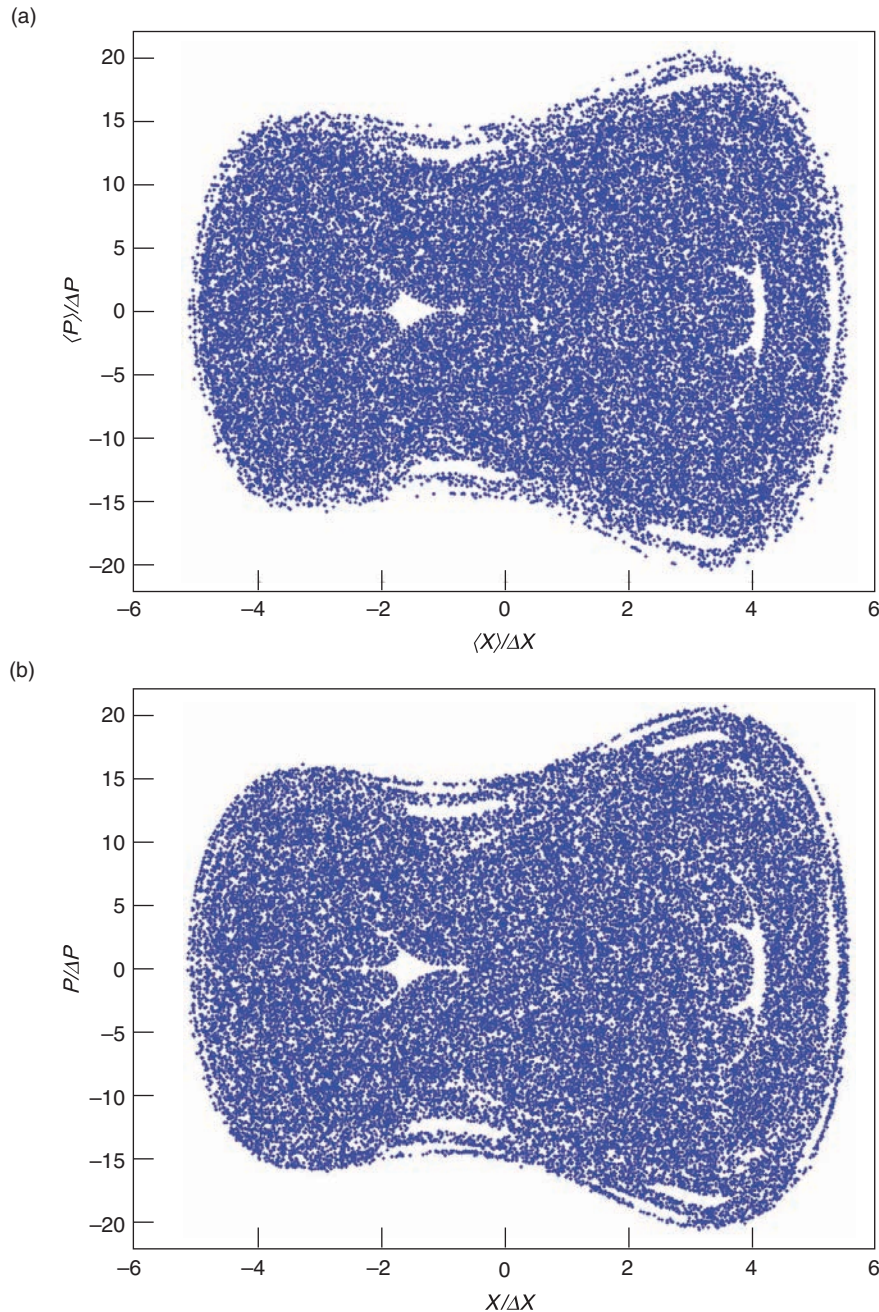


Figure 3. Stroboscopic Maps for the Quantum and Classical Duffing Oscillators

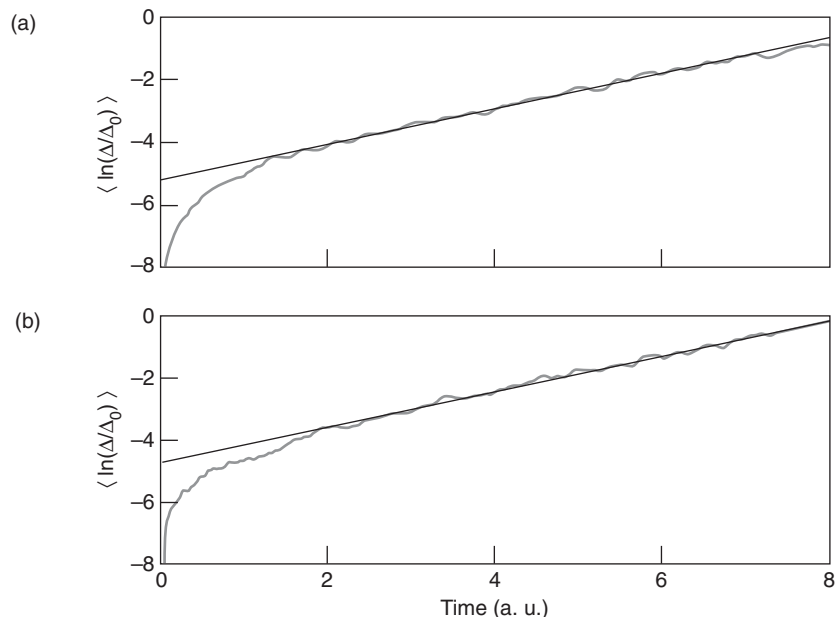
The results of the Duffing oscillator simulations are plotted as stroboscopic maps. (a) The map for the continuously observed quantum Duffing oscillator displays the centroids of the Wigner function at time intervals separated by the period of the driving force. This map is a pastiche from several different runs with different initial conditions, for a total duration of 39,000 periods of the temporal drive. (b) The map for the classical Duffing oscillator driven with a small amount of noise displays the calculated locations of particles in phase space at time intervals separated by the period of the driving force. The two maps are very similar. The quantum system under continuous measurement exhibits qualitatively the same chaotic behavior as the classical system driven with a small amount of noise. In these figures, $\Delta X = 33$ nm, and $\Delta P = 324$ pg nm/s.

and $p = 2.6$ pg $\mu\text{m/s}$. Those trajectories give a Lyapunov exponent of 5.6(1) per second, which is in agreement with the previous value.

Now we elaborate on the problem hinted at in the introduction. If observation realizes the classical world, do trees in remote forests fall quantum mechanically? Of course, the tongue-in-cheek answer is, “who knows?” At a deeper level, however, we note that even in a remote forest, trees continue to interact with the environment, and through this interaction, the components of the environment (reflected light, air molecules, and so on) acquire information about the system. According to unitarity, an important property of quantum mechanics, information can never be destroyed. The information that flowed into the environment must either return to its origin or stay somewhere in the environment—the decaying sound of the falling tree must yet record

Figure 4. Lyapunov Exponents for the Quantum and Classical Duffing Oscillators

In order to calculate the Lyapunov exponents, λ , for (a) a continuously observed quantum Duffing oscillator and (b) a classical Duffing oscillator driven with a small amount of noise, we plot against time the logarithm of the average separation of trajectories that begin very close together. The parameters defining the oscillator—the continuous-measurement strength in the quantum system and the noise in the classical system—are detailed on pages 119–120 of this article. The slope of the line drawn through the curves gives the Lyapunov exponent, which in both cases is $\lambda = 0.57(2)$. Also in both cases, $\Delta_0 = 33$ nm.



its presence faithfully, albeit perhaps only in a shaken leaf. And herein lies the key to understanding the unobserved: If a sufficiently motivated observer were to coax the information out of the environment, that action would become an act of continuous measurement of the current happenings even though actually performed in the future. But since the current state of affairs can't be influenced by what anyone does in the future, the behavior of the system at present cannot contradict anything that such a classical record could possibly postdict.

If the motion is not observed, no one knows which of the possible paths the object took, but the rest of the universe does record the path, which could, therefore, be considered as classical as any (Gell-Mann and Hartle 1993). All that happens when there is no observer is that our knowledge of the motion of the object is the result of averaging over all the possible trajectories. In that case, we are forced to describe the state of the system as being given by a probability distribution in phase space since we no longer know exactly where the system is as it evolves. This observation is, however, just as true for a (noisy) classical system as it is for a quantum system.

The Connection to the Theory of Decoherence

We can now explain how the analysis presented here relates to a standard approach to the quantum-to-classical transition often referred to as decoherence. The procedure employed in decoherence theory is to examine the behavior of the quantum system coupled to the environment by averaging over everything that happens to the environment. This procedure is equivalent to averaging over all the possible trajectories that the particle might have taken, as explained above. Thus decoherence gives the evolution of the probability density of the system when no one knows the actual trajectory. The relevant theoretical tools for understanding this process were first developed and applied in the 1950s and 1960s (Redfield 1957, Feynman and Vernon 1963), but more recent work (Hepp 1972, Zurek 1981, 1982, Caldeira and Leggett 1981, 1983a, 1983b,

Joos and Zeh 1985) was targeted at condensed-matter systems and a broader understanding of quantum measurement and quantum-classical correspondence. It was found that averaging over the environment or over the equivalent, unobserved, noisy classical system gives the same evolution (Habib et al. 1998). In this classical counterpart, different realizations of noise give rise to slightly different trajectories, and in a chaotic system, these trajectories diverge exponentially fast. As a result, probability distributions obtained by averaging over the noise tend to spread out very fast, and our knowledge of the system state is correspondingly reduced. In other words, discarding the information that is contained in the environment or, equivalently, the measurement record, as averaging over these data implies, leads to a rapid loss of information about the system. This increasing loss of information, characterized by a quantity called entropy, can then be used to study the phenomenon of chaos with varying degrees of rigor.

Averaging over the environment to produce classical probability distributions was, however, not completely satisfactory. Not only does this averaging procedure not allow us to calculate trajectory-based quantities, but it also restricts our predictions to those derivable by knowing only the probability densities at various times. But classical physics is much more powerful than that—it can predict the outcome of many “if ... then” scenarios. If I randomly throw a ball in some direction, the probability of it landing in any direction around me is the same, but if you see the ball north of me, you can predict with pretty good certainty that it won’t land south of me. In the classical world, such correlations are numerous and varied, and the measurement approach we have taken here completes our understanding of the quantum-to-classical transition by treating all correlations on an equal footing. It is easy to see, however, that if the continuous measurement approach has to get all the correlations right, it must per force get the decoherence of probability densities right!

The realization that continuous measurement was the key to understanding the quantum-to-classical transition has emerged only in the last decade. First introduced in a paper by Spiller and Ralph (1994), this idea was then mentioned again by Martin Schlautmann and Robert Graham (1995). Subsequently, the idea was developed in a collection of papers (Schack et al. 1995, Brun et al. 1996, Percival and Strunz 1998, Strunz and Percival 1998). However, the scientific community was slow to pick up on this work, possibly because the authors used a stochastic model referred to as quantum state diffusion, which may have obscured somewhat the measurement interpretation. In 2000, we published the results presented in this article, namely, analytic inequalities that determine when classical motion will be achieved for a general single-particle system, and showed that the correct Lyapunov exponent emerges (Bhattacharya et al. 2000). For this purpose, we used continuous position measurement, which is ever present in the everyday world and therefore the most natural one to consider. This accumulation of work now provides strong evidence that continuous observation supplies a natural and satisfactory explanation for the emergence of classical motion, including classical chaos, from quantum mechanics. In addition, such an analysis also makes clear that the specific measurement model is not important. Any environmental interaction that provides sufficient information about the location of the system in phase space will induce the transition in macroscopic systems. Recently, Andrew Scott and Gerard Milburn (2001) have analyzed the case of continuous joint measurement of position and momentum and of momentum alone, and they verified that classical dynamics emerges in the same way as described in Bhattacharya et al. (2000). ■

Further Reading

- Bhattacharya, T., S. Habib, and K. Jacobs. 2000. Continuous Quantum Measurement and the Emergence of Classical Chaos. *Phys. Rev. Lett.* **85**: 4852.
- Brun, T. A., I. C. Percival, and R. Schack. 1996. Quantum Chaos in Open Systems: A Quantum State Diffusion Analysis. *J. Phys. A* **29**: 2077.
- Caldeira, A. O., and A. J. Leggett. 1981. Influence of Dissipation on Quantum Tunneling in Macroscopic Systems. *Phys. Rev. Lett.* **46**: 211.
- . 1983a. Quantum Tunneling in a Dissipative System. *Ann. Phys. (N.Y.)* **149**: 374.
- . 1983b. Path Integral Approach to Quantum Brownian-Motion. *Physica A* **121**: 587.
- Carmichael, H. J. 1993. *An Open Systems Approach to Quantum Optics*. Berlin: Springer-Verlag.
- Davies, E. B. 1976. *Quantum Theory of Open Systems*. New York: Academic Press.
- DeWitt, B. S., and N. Graham, Eds. 1973. *The Many-Worlds Interpretation of Quantum Mechanics*. Princeton: Princeton University Press.
- Einstein, A. 1917. On the Quantum Theorem of Sommerfeld and Epstein. *Verh. Dtsch. Phys. Ges.* **19**: 434.
- Feynman, R. P., and F. L. Vernon. 1963. The Theory of a General Quantum System Interacting with a Linear Dissipative System. *Ann. Phys. (N. Y.)* **24**: 118.
- Gell-Mann, M., and J. B. Hartle. 1993. Classical Equations for Quantum Systems. *Phys. Rev. D* **47**: 3345.
- Ghose, S., P. M. Alsing, I. H. Deutsch, T. Bhattacharya, S. Habib, and K. Jacobs. 2002. Recovering Classical Dynamics from Coupled Quantum Systems Through Continuous Measurement. [Online]: <http://eprints.lanl.gov/quant-ph/0208064>.
- Habib, S., K. Shizume, and W. H. Zurek. 1998. Decoherence, Chaos, and the Correspondence Principle. *Phys. Rev. Lett.* **80**: 4361.
- Habib, S., K. Jacobs, H. Mabuchi, R. Ryne, K. Shizume, and B. Sundaram. 2002. The Quantum-Classical Transition in Nonlinear Dynamical Systems. *Phys. Rev. Lett.* **88**: 040402.
- Hepp, K. 1972. Quantum Theory of Measurement and Macroscopic Observables. *Helv. Phys. Acta* **45**: 237.
- Joos, E., and H. D. Zeh. 1985. The Emergence of Classical Properties through Interaction with the Environment. *Z. Phys. B* **59**: 223.
- Kraus, K. 1983. *States, Effects, and Operations: Fundamental Notions of Quantum Theory*. Berlin: Springer-Verlag.
- Landau, L. D., and E. M. Lifshitz. 1965. *Quantum Mechanics: Non-relativistic Theory*. New York: Pergamon Press.
- . 1980. *Statistical Physics*. New York: Pergamon Press.
- Omnès, R. 1994. *The Interpretation of Quantum Mechanics*. Princeton: Princeton University Press.
- Percival, I. C., and W. T. Strunz. 1998. Classical Dynamics of Quantum Localization. *J. Phys. A* **31**: 1815.
- Poincaré, H. *New Methods of Celestial Mechanics*. 1992. Edited by D. L. Goroff. New York: Springer.
- Redfield, A. G. 1957. Theory of Relaxation Processes in Advances in Magnetic Resonance. *IBM J. Res. Dev.* **1**: 19.
- Schack, R., and C. M. Caves. 1999. Classical Model for Bulk-Ensemble NMR Quantum Computation. *Phys. Rev. A* **60**: 4354.
- Schack, R., T. A. Brun, and I. C. Percival. 1995. Quantum State Diffusion, Localization and Computation. *J. Phys. A* **28**: 5401.
- Schlautmann, M., and R. Graham. 1995. Measurement Trajectories of Chaotic Quantum Systems. *Phys. Rev. E* **52**: 340.
- Scott, A. J., and G. J. Milburn. 2001. Quantum Nonlinear Dynamics of Continuously Measured Systems. *Phys. Rev. A* **63**: 042101.
- Spiller, T. P., and J. F. Ralph. 1994. The Emergence of Chaos in an Open Quantum System. *Phys. Lett. A* **194**: 235.
- Strunz, W. T., and I. C. Percival. 1998. Classical Mechanics from Quantum State Diffusion—A Phase Space Approach. *J. Phys. A* **31**: 1801.
- Zurek, W. H. 1981. Pointer Basis of Quantum Apparatus: Into What Mixture Does the Wave Packet Collapse? *Phys. Rev. D* **24**: 1516.
- . 1982. Environment-Induced Super-selection Rules. *Phys. Rev. D* **26**: 1862.
- . 1991. Decoherence and the Transition to Classical. *Phys. Today* **44**: 36.

Tanmoy Bhattacharya graduated from the Indian Institute of Technology in Kharagpur, India, with a master's degree in physics in 1984. He received a



Ph.D. in physics from the Tata Institute of Fundamental Research in Bombay, India, in 1989. Tanmoy worked as a postdoctoral researcher at Brookhaven National Laboratory, at the Centre d'Énergie Atomique in Saclay, France, and at

Los Alamos National Laboratory before becoming a staff member in the Theoretical Division at Los Alamos in 1995. Over the years, Tanmoy's research activities have diversified. Having started with an interest in the structure and interactions of elementary particles, Tanmoy has become interested in phylogenetic problems in viral evolution and fundamentals of quantum mechanics. He is currently working on the interface of quantum and classical systems.

Kurt Jacobs received a bachelor of science degree from the University of Auckland, New Zealand, in 1993 and a master's degree in physics from the same institution in 1995. Three years later, Kurt received his Ph.D. in physics from Imperial College, London University. Since that time, Kurt has been working as a postdoctoral fellow in the Elementary Particles and Field Theory



Group of the Theoretical Division at Los Alamos National Laboratory. His work has been mainly in the fields of quantum measurement theory, the quantum-to-classical transition, and quantum information theory.

Salman Habib received his undergraduate degree from the Indian Institute of Technology in Delhi, India, and his Ph.D. from the University of Maryland, College Park, in 1988. He then held postdoctoral fellowships at the University of British Columbia, Vancouver, and at Los Alamos National Laboratory. In 1994, he became a staff member in the Theoretical Division at Los Alamos. The

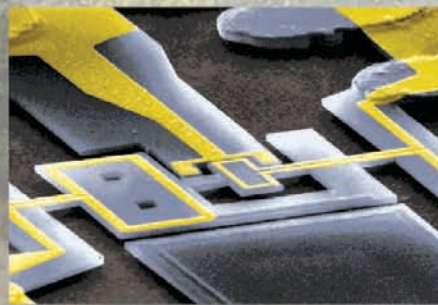
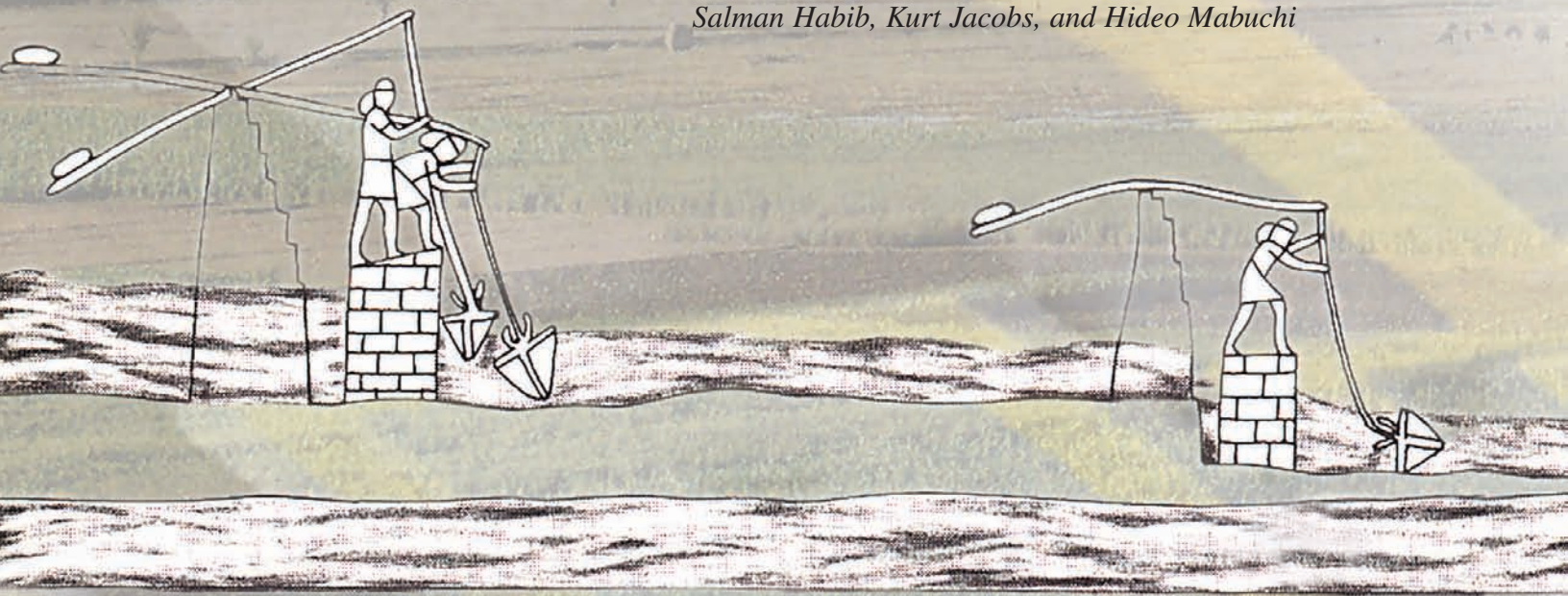


central theme of Salman's research has been the study of dynamical systems covering areas that range from classical and quantum chaos to the dynamics of the Universe. In the past decade, Salman has contributed to elucidating the nature of the quantum-to-classical transition and worked on the problem of controlling quantum dynamical systems. His recent work focuses strongly on the interface between theory and experiment.


Quantum Feedback Control

How can we control quantum systems without disturbing them?

Salman Habib, Kurt Jacobs, and Hideo Mabuchi



The nanomechanical electrometer shown here was built in Michael Roukes' group at Caltech. It has a demonstrated sensitivity below a single electron charge per unit bandwidth and should ultimately reach sensitivities of the order of parts per million. Its operation is based on the movement of a torsional resonator that carries a detection electrode placed in an external magnetic field. The gate electrode is seen on one side of the resonator.



Ever since Niels Bohr's first attempt at understanding the hydrogen atom, the fundamental cautionary lesson of quantum mechanics has been driven home time after time: Processes in the microworld transpire according to laws and principles that directly contradict those governing the macroworld of human experience. This radical shift in understanding is now almost a century old and has been definitively confirmed by numerous experiments. It might seem likely that the strange behaviors of quantum systems would be familiar by now and practical devices harnessing those behaviors would be commonplace. For the most part, however, we have remained mere spectators of the microphysical realm, where quantum mechanics holds sway, being forced to observe naturally occurring phenomena rather than being able to control and manipulate them. In the coming decade, however, this situation may be reversed.

Recent advances in quantum and atomic optics and condensed matter physics are providing tools to engineer practical quantum devices and perhaps even modestly complex networks of these devices. Quantum information processing, precision measurement, and development of ultrasensitive sensors are driving the present development of quantum technologies. If quantum technologies are ever to achieve the complexity of classically engineered systems such as jet aircraft and the Internet, a quantum analog of classical feedback control must be developed, since feedback control is at the heart of the stability and predictability underlying complex engineered systems.

Along these lines, recent theoretical results on error correction in quantum computation and on the dynamics of open quantum systems may be viewed as first steps in developing a theoretical formalism for practical quantum feedback control (see the articles "Introduction to Quantum Error Correction" on page 188 and "Realizing a Noiseless Subsystem" on page 260). Indeed, feedback control represents a promising new approach to mitigating quantum noise and decoherence in both quantum computation and precision measurement. If we are to apply the concepts and methods of feedback control theory to quantum dynamical systems, we must not only extend classical control concepts to new regimes but also analyze quantum measurement in a way that is useful for control systems.

The Evolution of Control Theory

Controlling natural phenomena through macroscopic engineering goes back thousands of years. Consider for a moment the ingenious ways in which early human civilizations controlled irrigation. In Mesopotamia (2000 BC), where rainfall was poor and the Tigris and Euphrates Rivers were the main sources of water, engineers constructed an elaborate canal system with many diversion dams (see the drawing to the left). In that system, the Euphrates served as a source and the Tigris as a drain. In a similar vein, the ancient Egyptians used water from the Nile and thereby allowed their civilization to flourish. On a smaller scale, machines using feedback control were developed in the

Greco-Roman period, and methods for the automatic operation of windmills date back to the Middle Ages.

Perhaps the best-known example of feedback control in the industrial era is the Watt governor, which stabilizes steam engine speeds under fluctuating loads. James Clerk Maxwell provided the first dynamical analysis of this system based on differential equations. His work, which was published in 1868, founded the field of mathematics now known as control theory. In the early part of the 20th century, the idea of self-regulating machinery continued to be pushed in various directions, notably in electronic amplification. Control concepts were further developed for industrial, navigational, and military applications.

After World War II, control systems progressed to a new level of complexity. Up until that time, feedback control systems had been largely single loop, taking the feedback signal from one point and connecting the correction signal to a different point. Multiloop control systems and more sophisticated feedback techniques emerged from progress in optimization theory and dynamical systems theory, as well as from the advent of digital computers.

After 1960, there emerged what is often referred to as "modern" (as opposed to "classical") control theory (Brogan 1990, Zhou et al. 1996), which emphasizes optimization of cost and performance. For the same control goals, it is clear that not all control strategies will be equally effective in terms of cost and performance. Determining the best strategy defines the problem of optimal control; however, optimal algorithms are often unstable to variations in system

parameters and the external environment. Theorists then turned to ensuring performance bounds in the presence of uncertainty. This work resulted in the theory of “robust” control (Zhou et al. 1996). Noise in the inputs, extrinsic disturbances in the system under control, measurement errors, and modeling inadequacies—all can render control systems less effective or, in some cases, even lead to catastrophic failures. The role of robust control is to maintain adequate stability and other performance margins given the uncertainties mentioned earlier.

Classical Control Systems

Formally speaking, a control system consists of a dynamical system interacting with a controller, a device that influences the state of the dynamical system toward some desired end. The objective may be to regulate the flow of an industrial process, money, energy, information, and so on. In a “closed-loop,” or feedback, control system, the controller uses outputs from the dynamical system to monitor and influence its interaction with that dynamical system. For a linear dynamical system, for example, such a situation could be described by the following equation:

$$dx = Axdt + BdW + Cu , \tag{1}$$

where \mathbf{x} is a vector describing the state of the system, dW is a vector of Gaussian noise sources, and \mathbf{u} is the vector of inputs determined by the controller. The matrix A gives the system’s deterministic motion, and B and C describe, respectively, how the noise and input vectors are coupled into the system. A separate equation, namely,

$$dy = Hxdt + RdV , \tag{2}$$

describes the continuous measurement

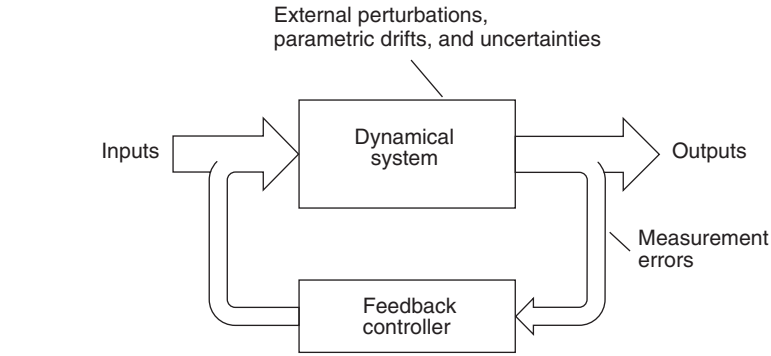


Figure 1. Classical Feedback Control

The classical dynamical system to be controlled has a set of input variables, which are processed by the system dynamics into a set of output variables. Some fraction of the set of input and output variables (possibly different for each case) is available for hookup to the controller. The controller has to perform in the presence of external fluctuations—that is, uncertainties and drifts in the parameters describing the dynamical system—and measurement errors.

of system outputs by the controller. In each small time interval dt , the controller obtains the measurement result dy . That result is directly related to the true state of the system by some linear transformation H , but it also includes a Gaussian noise process V , which serves to represent imperfections in the measurement.

Examples of control systems can be found in many applications. For instance, servomechanisms are control systems that use small control inputs to produce changes in large mechanical systems. In effect, the larger systems are “slaved” to the output of the servomechanisms (for example, liquid levels in reservoirs are controlled by float valves). Feedback circuits are used in ingenious ways in electronic amplification to manipulate input and output impedances and to improve the linearity, distortion, and frequency bandwidth of the output signal relative to the input signal.

In an “open-loop” control system, the controller does not monitor the output of the dynamical system. A dynamical model for the system is assumed, and control is applied with the idea that the desired outcome will actually be achieved. Open-loop

strategies are useful in situations in which the system dynamics are known precisely and vary only slowly. Processes with long measurement dead times are sometimes better suited to open-loop control methods than to feedback methods. Open-loop control strategies are applied in situations as diverse as the maximization of returns from financial investments, optimal determination of aircraft flight paths, and controlled dissociation of molecules.

Figure 1 shows how to implement closed-loop control for a dynamical system. One must be able to measure some of the dynamical variables of the system under control (the outputs) and use them to influence some other variables (the inputs). In other words, given the output variables, the controller implements a particular control strategy to influence the state of the dynamical system by appropriately varying the inputs. Robust controllers take into account variations in system parameters and fluctuations from the external environment to produce control strategies with guaranteed stability bounds.

Control systems can involve many different interacting physical systems

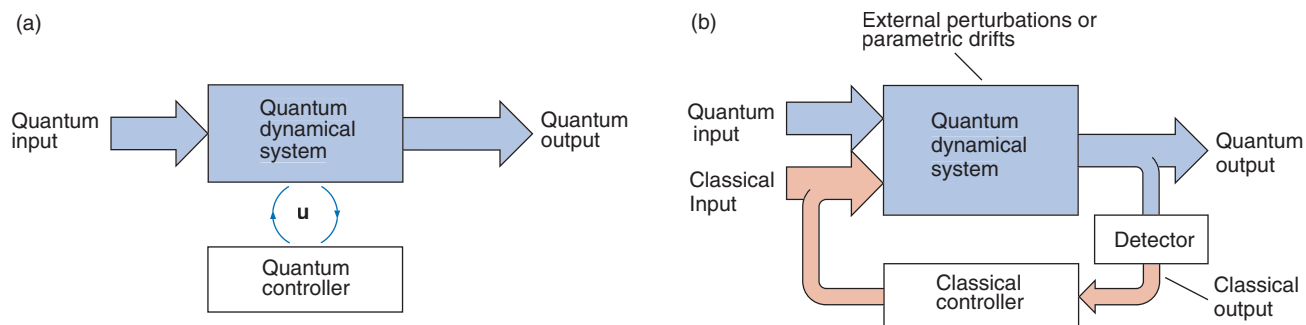


Figure 2. Directly and Indirectly Coupled Quantum Feedback

(a) Both the dynamical system and the controller are quantum systems coupled through a unitary interaction. A quantum variable is coupled to the quantum controller, and a quantum input path from the controller goes directly back to the quantum system. (b) A quantum dynamical system can be viewed as having two sets of inputs, one relating to the variation in the classical parameters describing the Hamiltonian and the other representing fully quantum inputs. Similarly, the output channel can be divided into a quantum and a classical channel. The classical channel is, in fact, a piece of the quantum channel that has become classical after observation. The controller analyzes the classical record to form an estimate of the dynamical system's state and uses this information to implement the appropriate control.

with a large number of sequential, parallel, and nested control loops that are both open and closed. For example, closed- and open-loop strategies can be combined as in the fast closed-loop systems used to stabilize the slower, inherently unstable open-loop dynamics of modern fighter aircraft.

Developing Control in Quantum Systems

The general picture of control systems outlined in the previous section appears to be extendable to quantum systems. Certainly, open-loop control problems are conceptually straightforward in the quantum context. One begins with the time evolution operator of the quantum system—the Schrödinger equation for the wave function, the Liouville equation for the density matrix, or more complicated dynamical evolution equations for the density matrix characterizing a system coupled to an environment. A theory for time-dependent variations in the evolution operator is then developed in such a way that the wave function or the density operator at some time is close to some target

value. This target value does not have to be unique, nor in fact is the time evolution to that value unique. The approach just outlined applies equally well to classical probabilistic evolutions: Although quantum and classical systems are dynamically distinct, the principles for open-loop control are in fact very similar.

Controlling chemical reactions by laser-produced electromagnetic fields that are time dependent is a well-known open-loop quantum control problem. In the frequency-resolved approach to control, the quantum interference between different evolutionary paths is being manipulated; in the time-resolved approach, the dynamics of wave packets produced by ultrafast laser pulses leads to control. For some specific control of the chemical reactions, one can optimize the temporal and spectral structure of those laser pulses (Shi et al. 1988).

The fundamental differences between classical and quantum systems become real issues, however, in the field of closed-loop control. Quantum systems can have two distinct types of feedback control: directly and indirectly coupled quantum feedback (see Figure 2). As illustrated in Figure 2(a),

in a system with directly coupled quantum feedback, a quantum variable of the system is coupled to the quantum controller, and a quantum input path from the controller goes directly back to the quantum system. When the quantum feedback is indirect, as shown in Figure 2(b), the quantum dynamical system under control is an observed system. It therefore generates a classical output, also known as the measurement record, which the controller may analyze to provide a best estimate of the original quantum state of the system. The controller then feeds back a classical signal to vary parameters in the quantum evolution operator in accord with the chosen control strategy. Hybrid couplings using both direct and indirect quantum feedback channels are easy to envisage: The channel from the system output to the controller input may be directly coupled whereas the channel from the controller output to the system input may be coupled indirectly through a classical path.

In both classical and quantum contexts, the main goal of closed-loop control is to enhance system performance in the presence of noise from both the environment and the

uncertainty in the system parameters. To limit the effects of noise, the controller must perform an irreversible operation. Noise generates a large set of undesirable evolutions, and the controller's task is to map this large set to a much smaller one of more desirable evolutions. Mapping from the larger to the smaller set is by definition irreversible. In other words, noise is a source of entropy for the system. To control the system, the controller must extract the entropy from the system under control and put it somewhere else. The controller must therefore have enough degrees of freedom to respond conditionally upon the noise realization. In indirect quantum feedback control, the measurement process, coupled with the conditional response of the controller, is the source of entropy reduction. In direct quantum feedback control, the evolution of the system is fully unitary, or quantum mechanical. The quantum controller provides a large Hilbert space of quantum mechanical states. That is precisely where the entropy generated by the noise may be put (or where the history of the effect of the noise on the system may be stored). The quantum controller then reacts conditionally to this quantum record, keeping the entropy of the quantum dynamical system low, while the entropy of the storage location grows continually.

Inherent Noise Generation in Quantum Feedback Control

Unlike classical systems, quantum systems may be easily disturbed when information about them is extracted. Measurement disturbs a quantum system through the following intrinsic property of quantum mechanics: Obtaining accurate knowledge about one observable of a quantum system necessarily limits the information about an observable conjugate to the

first. For example, particle position and momentum are conjugate observables, and the uncertainties inherent in the knowledge of both are codified by the famous Heisenberg uncertainty relation. If the chosen feedback-control strategy involves measurement, one must take into account the effects of the measurement on the evolution of the quantum system. A generally applicable model for including those effects is that of a continuous quantum measurement. This model was developed for quantum optics (Carmichael 1993), a field in which such measurements have been realized experimentally, and it was also derived in the mathematical physics literature with the help of more abstract reasoning (Barchielli 1993). In this volume, the model of a continuous quantum measurement is presented in the article "The Emergence of Classical Dynamics in a Quantum World" on page 110.

Quantum measurements may introduce unwanted noise in three more-or-less distinct ways. First, one may measure an observable conjugate to the real variable of interest and thereby introduce more uncertainty in the latter variable. More generally, one may attempt to obtain information inconsistent with the state under control. For example, to preserve a state that is the superposition of two position states, position measurements must be avoided because they will destroy the superposition. Thus, in quantum mechanics, the type of measurement chosen must be consistent with the control objectives. This condition is unnecessary in classical feedback control. Second, if trying to control the values of observables (Doherty et al. 2000), one must consider that the time evolutions of different observables necessarily affect each other over time. Observables whose values are uncertain at one time will cause other observables (perhaps more accurately known) to become uncer-

tain at a later time. For example, a very accurate measurement of the particle position at one time introduces uncertainty into the value of the particle momentum. Because the value of momentum determines the position of the particle at a later time, the momentum uncertainty makes the future position of the particle more uncertain, hence introducing noise into the quantity that is being measured. This mechanism for introducing noise is usually referred to as the back action of a quantum measurement.

The third kind of noise involves the randomness of the measurement results. Because the state of the observed system after a measurement depends upon the outcome of the measurement, the more the result fluctuates, the more noise there is in the evolution of the system. For classical measurements, fluctuations in measurement results cannot be any more than the entropy of the system before measurement; that is, the measurement does not introduce any additional noise into the system. In quantum mechanics, however, even if the system state is known precisely, one can still make measurements that change the state in a random way, thereby actually injecting noise into the system. This observation is particularly relevant when the overall state of the system, rather than a specific observable, is being controlled. The situation is further complicated by the fact that, for certain classes of measurements, there is actually a tradeoff between the noise injected by the measurement and the information gained by the observer (Doherty et al. 2001). As a result, designing measurement strategies is far from being a trivial activity.

Strategies for Quantum Feedback Control

The differences between classical and quantum measurements profoundly

affect the design of feedback control algorithms. A classical controller extracts as much information from the system as possible. In quantum control, irreducible disturbances are inherent to any measurement, and therefore the measurement strategy becomes a significant part of the feedback algorithm. For example, just as the inputs to the system change with time, the measurements too may need to be varied with time so that the best control should be achieved.

Adaptive measurement, or altering the measurement as it proceeds, was first introduced by Howard Wiseman (1995), not for control but for accuracy. The result was a more accurate measurement of some aspect of the quantum state. Nevertheless, this approach has a unique bearing on quantum feedback control algorithms. Knowing that quantum measurements can disturb the state being measured, one may want to start a continuous measurement process by measuring in a way that is not necessarily optimal but is sufficiently weak to cause minimal disturbance to the aspect of interest. As the measurement proceeds, one uses the continuously obtained information about the state to make the measurement increasingly close to optimal.

For example, consider measuring the oscillation amplitude of a harmonic oscillator when the phase of the oscillation is unknown but the oscillator is known to be in an amplitude-squeezed state; that is, the uncertainty in amplitude or energy is much smaller than the uncertainty in phase, the conjugate variable (see Figure 3). In this case, an accurate measurement of amplitude is given by a measurement of position at the moment when the particle is at its maximum spatial extent, or maximum distance from $x = 0$. On the other hand, at the moment when the particle has the most momentum (at position $x = 0$), the ideal quantity to

measure is momentum. Thus, for a continuous measurement of the oscillation amplitude, a linear combination of position and momentum should be measured and the relative weighting of those two variables should be allowed to oscillate in time. However, without knowing the mean phase of oscillation, one cannot know which variable should have the most weighting in the measurement at what time. Using an adaptive measurement procedure, one can start by assuming the oscillator to have a particular phase and then adjust the relative weights of position and momentum to more desirable values as information about the phase is obtained.

Applications of Quantum Control

Atomic optics is one field in which it should be possible to test quantum

feedback control in the near future.

It has already been demonstrated that a single atom can be trapped inside an ultralow-loss optical cavity (mirror reflectivity is $R = 0.9999984$ in experiments at Caltech) in the strong-coupling quantum regime (Mabuchi et al. 1999). Figure 4 illustrates the experimental setup used at Caltech. The strong coupling occurs between the atom and the radiation field in the cavity and is proportional to the induced atomic dipole moment and the single-photon cavity field. Continuous measurements and real-time feedback could be used to cool such an atom to the “ground” state of the quantized mechanical potential produced by several photons in the cavity. The average number of photons circulating inside such a cavity can be kept very low (from 1 to 10 photons) if one uses a weak driving laser that barely balances the slow rate at which individual photons leak out. If the cavity mode volume is

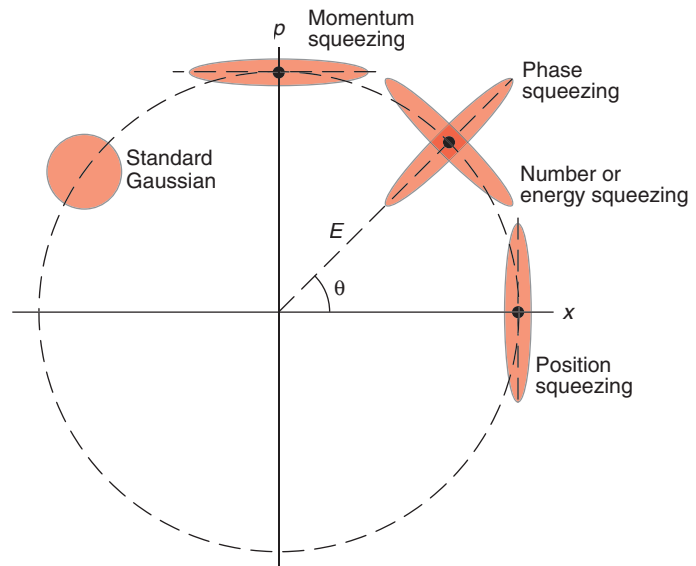


Figure 3. “Squeezed” States for a Harmonic Oscillator

Squeezing may be illustrated by considering phase-space plots of a Gaussian wave function. For a standard Gaussian state, the uncertainties in the x - and p -direction are equal, and the uncertainty ellipse takes the shape of a circle, provided appropriate position and momentum scalings are made. When states are “squeezed,” the area of the uncertainty ellipse remains constant, but the ellipse is rotated and squeezed as shown. Squeezing momentum, for example, means reducing the uncertainty in momentum. The constant energy surface is the dashed circle, and the position on the circle can be specified by the angle. Squeezing phase and energy again refers to changes in shape of the uncertainty ellipse for the wave function.

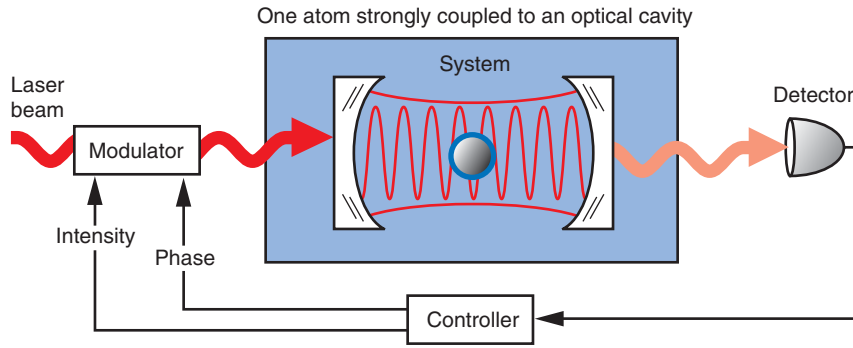


Figure 4. Quantum Feedback in a Cavity Quantum Electrodynamics Application

The dynamics between the atom and the photon field in the cavity can be modified by continuous measurement of the light transmitted through the cavity (which bears information about the evolving system state) and by continuous adjustment of the amplitude/phase of the driving laser in a manner that depends on the measurement results. Control objectives of fundamental interest include active cooling of the motion of an individual atom, feedback-stabilized quantum state synthesis, and active focusing of atomic beams for applications such as direct-write lithography.

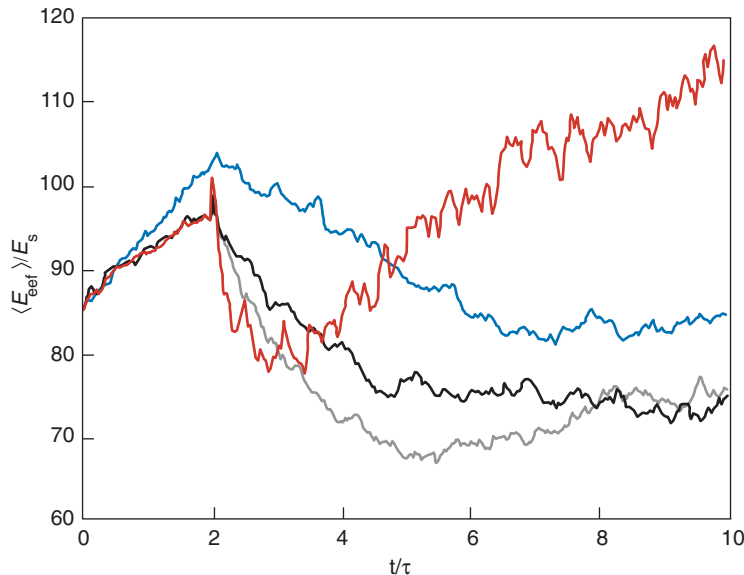


Figure 5. Simulating a Feedback Algorithm to Cool Atomic Motion in an Optical Cavity

In this simulated experiment, the light forms an effective sinusoidal potential for the atom, and the controller switches this potential between a high and a low value (separated by some ΔV) to cool the atomic motion. In this simulation, the feedback is turned on at $t = 2$, and the expected value of the atomic motion energy is plotted here as a function of time for four different values of ΔV . Although these results are still preliminary, they indicate that the effectiveness of the feedback algorithm is highly dependent on ΔV .

sufficiently small, just a few photons can give rise to dipole (alternating-current Stark shift) forces that are strong enough to bind an atom near a local maximum of the optical field distribution. At the same time, the atomic motion can be monitored in real time by phase-sensitive measurements of the light leaking out of the cavity. To a degree determined by the fidelity of these phase measurements, the information gained can be used continually to adjust the strength of the driving laser (and hence the depth of the optical potential) in a manner that tends to remove kinetic energy from the motion of the atomic center of mass.

In order to perform such a task in real time, however, it is essential to develop approximate techniques for continuously estimating the state of the atomic motion. Approximations are needed because integrating a stochastic conditioned-evolution equation to obtain a continuous estimate of the density matrix is far too complex a task to be performed in real time. While this experiment remains to be carried out, we have developed an approximate estimation algorithm¹ and used it in combination with an experimentally realizable feedback algorithm (see Figure 5).

Feedback cooling ideas can also be applied to condensed-matter systems. Some of our recent calculations predict that feedback control can be used to cool a nanoresonator below the limits set by refrigeration. This method would reduce thermal fluctuations to approximately the quantum energy level spacing of the resonator. These findings are important because nanoscale devices are interesting from a more fundamental perspective than merely sensing and actuation applications. Provided they can be cooled to

¹ This algorithm is described in a yet unpublished paper by Salman Habib, Kurt Jacobs, Hideo Mabuchi, and Daniel Steck.

sufficiently low temperatures, low-loss nanomechanical resonators would be excellent candidates for the first observation of quantum dynamics in mechanical mesoscopic systems. Yet, as mentioned above, in order to achieve this goal, we must reduce thermal fluctuations to approximately the quantum energy level spacing of the resonator, a task which requires temperatures in the range of millikelvins.

To cool the position coordinate of the nanoresonator, one needs a suitable scheme for continuous position measurement. One practical method of performing a continuous measurement of a nanoresonator's position is to use a single-electron transistor (SET)—see Figure 6. To make the measurement, one locates the resonator next to the central island of the SET. When the resonator is charged and the SET is biased so that current flows through it, changes in the resonator's position modify the energy of the central island, which produces changes in the SET current. The current therefore provides a continuous measurement of the position of the resonator, a requirement for implementing a linear feedback cooling algorithm. A feedback force can be applied to the resonator by varying the voltage on a “feedback electrode,” which is capacitively coupled to the resonator (see Figure 6). The applied voltage is adjusted so as to damp the amplitude of oscillation.

Experiments on nanomechanical oscillators observed with SETs currently start at temperatures near 100 millikelvins. These oscillators have fundamental frequencies f_0 on the order of 1 to 100 megahertz. As a concrete example, consider a practical oscillator with $f_0 = 10$ megahertz, a length of 2 micrometers, and the other two dimensions on the order of 100 nanometers. The effective mass of such an oscillator is roughly 10^{-19} kilograms. An achievable quality factor, Q , is about 10^4 . In order to

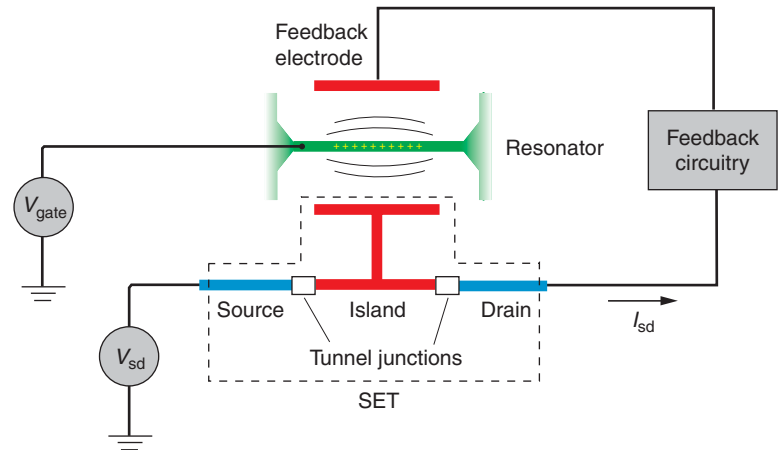


Figure 6. Cooling a Nanomechanical Resonator

This schematic diagram illustrates a concept for cooling a nanomechanical resonator to millikelvin temperatures, at which we can possibly observe quantum dynamics. An SET measures the position of the resonator, and a feedback mechanism damps (cools) the resonator's motion. The resonator, which is charged by the voltage source V_{gate} acts as the SET gate electrode. The resonator is also capacitively coupled to the SET island (red) and feedback electrode. As it moves back and forth relative to the SET island, the current I_{sd} flowing through the SET changes. Information about the changing current is used by the feedback circuitry to charge the feedback electrode. A force is generated that damps the resonator's oscillations.

observe discrete quantum passage from one oscillator energy level to another, the thermal energy should be on the order of the level spacing, that is, $k_B T \sim hf_0$, which corresponds to an effective temperature $T = .24$ millikelvin. Habib, Jacobs, Asa Hopkins, and Keith Schwab have shown that feedback cooling applied to this system at an initial temperature $T = 100$ millikelvins can yield a final temperature of $T = 0.35$ millikelvin. At this temperature, the aggregate occupation number lies between zero, the ground state, and one, the first excited state of the nanomechanical resonator. In other words, the system is cold enough to allow observation of quantum “jumps.” Although our calculations are based on certain idealized assumptions, those assumptions are close enough to reality that experimentalists can hope to achieve similar results.

Another, seemingly paradoxical, application of quantum feedback control techniques might be in sup-

pressing quantum dynamical effects such as tunneling. A classical memory device can be viewed as a two-state system with the two states separated by a finite energy barrier. At low temperatures, there is a finite probability of coherent or incoherent tunneling from one minimum to the other. Tunneling generates random memory errors, but continuous measurement, coupled with feedback, can suppress it. One such scheme is described and demonstrated in Andrew Doherty et al. (2000). The Hamiltonian for the double well is taken to be

$$H = \frac{1}{2} p^2 - Ax^2 + Bx^4, \quad (3)$$

where x and p are dimensionless position and momentum. Choosing $A = 2$ and $B = 1/9$ puts the minima of the wells at ± 3 and gives a barrier height of approximately 13.5. The controller is allowed to continuously observe the position of the particle and to apply a

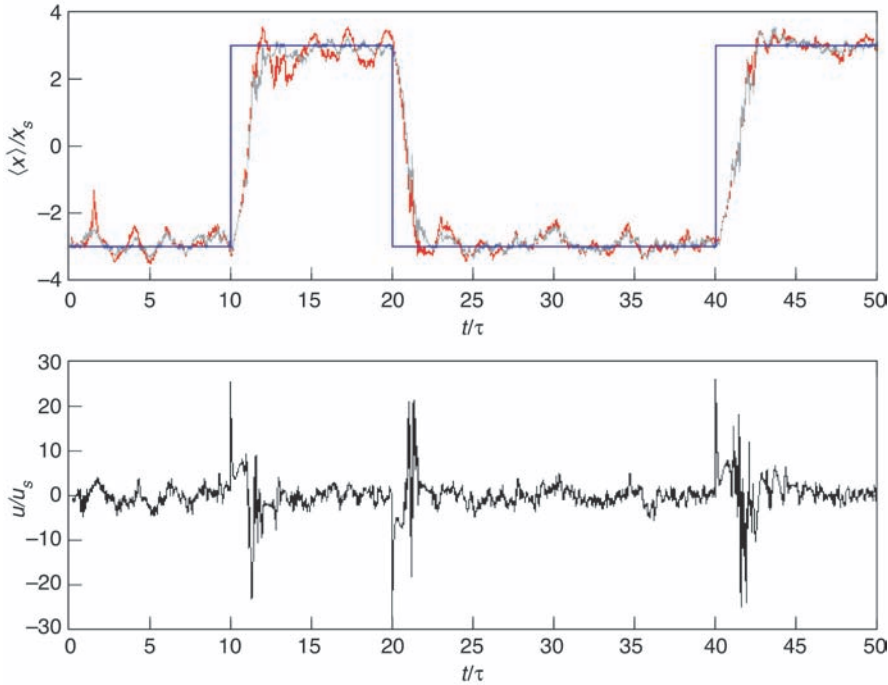


Figure 7. Particle in a Double Well Controlled by an Estimation-Feedback Scheme

(a) Shown here, as a function of time, are the target position (blue line), the “true” mean position (red line) obtained with the stochastic master equation (in which the measurement strength k equals 0.3 and the thermal heating rate β equals 0.1), and the position obtained with the Gaussian estimator (gray line). (b) The control strength (size of applied force) is shown as a function of time.

linear force in addition to the “double-well” potential already present. The continuous observation is described by the equation

$$dq = \langle x \rangle dt + \frac{dV}{\sqrt{8k}}, \quad (4)$$

where dq is the measurement result in the time interval dt and k is a constant characterizing the accuracy, or strength, of the measurement. The system is also driven by a thermal heat bath in the high-temperature limit. The effect of that bath is, in fact, the same as that of a continuous quantum measurement of position that ignores the measurement result. When the bath is described in this way, it is the strength of the fictitious measurement that gives the rate of thermal heating, and we will denote this constant by β .

Integrating a stochastic master equation gives the observer’s state of knowl-

edge as a result of the continuous measurement. However, since this is a differential equation for the density matrix of the single particle, it is numerically expensive to integrate. For practical purposes, one requires a simplified means for calculating a state estimate. To achieve this goal, we note that, as a result of the continuous observation, even though the dynamics are nonlinear, the density matrix remains approximately Gaussian. When a Gaussian approximation is used, the stochastic master equation reduces to a set of five equations (for all the moments of x and p up to quadratic order), and so it provides us with a practical method for obtaining a continuous state estimate. In practice, this Gaussian estimator can be shown to work quite well; that is, mean values from the approximate estimator agree very well with mean values derived from exact numerical solutions of the stochas-

tic master equation—see Figure 7(a).

In addition to a state estimation procedure, we also require a feedback algorithm. If the system were linear, one could apply the optimal techniques of modern control theory to find a feedback algorithm. Because attempting an optimal control solution for the full nonlinear problem is computationally intractable, the idea is to linearize the system dynamics around the present estimate of the state with the further assumption that the probability density, conditioned on the measurement record, remains Gaussian. As long as position measurements are sufficiently strong, this last condition is satisfied. The importance of this condition is twofold: Having a Gaussian approximation does not only mean that a small number of moments (five) are needed to describe the distribution but also that the quantum propagator is very close to the classical propagator at each time step (for exactly Gaussian states, the two are identical), and hence techniques borrowed from classical control have an excellent chance of working. The control can fail if the measurement is too weak to maintain a localized Gaussian distribution or if it is too strong. In the latter case, the state is Gaussian, but the measurement noise is too large.

The Gaussian state estimate is now used to set the value of the feedback term in the Hamiltonian (the sign and the magnitude of the coefficient of the linear feedback term in the potential). By choosing appropriate strengths for the measurement and the feedback strength, one can show that the feedback scheme is effective in controlling whether the particle is in the desired minimum—see Figure 7(b). For this plot, the measurement strength is $k = 0.3$, and the thermal heating rate is $d\langle E \rangle / dt = \beta = 0.1$.

This scheme has limitations arising from unwanted heating due to the measurement. Although some of the

heating derives directly from having to keep the state close to Gaussian, a more general limitation also contributes to heating: The measurement must be sufficiently strong to provide enough information for control to be effective. Developing new estimation and feedback schemes that can reduce the measurement-induced heating rate is an important area for future research.

Outlook for the Future

Most likely, ideas in quantum feedback control will first be tested in condensed matter physics and in quantum and atomic optics. Experiments in atomic optics have already furnished the cleanest tests and demonstrations of quantum mechanics in the last several decades. These include violations of the Bell inequalities, quantum teleportation, quantum state tomography, quantum cryptography, and single-atom interference. The ability to compare experimental results with precise theoretical benchmarks is a hallmark of these tests. As these experiments become increasingly sophisticated and complex, one can envisage a passage from “toy” demonstrations to real applications such as feedback control. The more strongly coupled systems of condensed matter physics are less amenable to accurate theoretical prediction. Nevertheless, experiments are becoming comparable in quality to early atomic optics experiments, and the time is ripe for active interaction between these two fields: Theoretical development in quantum optics, such as continuous measurement and quantum control, can be taken over to condensed matter contexts, most notably in nanotechnology. As the size of the smallest structures that can be fabricated by lithographic techniques decreases, the need for quantum mechanics becomes inevitable. Since lithography is the

only way we know to create very complex systems at reasonable cost, it follows that a fundamental and predictive understanding of quantum dynamics applicable to these systems (whether coherent or incoherent) will be required. It is also clear that, for these systems to be designable and to function reliably in an engineering sense, further development of quantum control theory will be necessary.

From a “more algorithmic” perspective, the Holy Grail is the development of optimal and robust control algorithms that are generally applicable. So far, apart from the trivial case in which the system dynamics are linear and the measurement strategy is considered fixed (Doherty and Jacobs 1999), no such optimal algorithms have been found for quantum feedback control. In classical control theory, optimal and robust control algorithms exist for linear systems, but only very few for nonlinear systems despite the best effort of control theorists in the past few decades. Nonlinear classical optimal control is a very difficult problem indeed, and probably intractable in most cases. Systematic numerical search algorithms for optimal strategies exist, but these also become intractable for systems of reasonable size. Because the dynamics of noisy and measured quantum systems is inherently nonlinear, the quantum control problem may also be intractable (Doherty et al. 2000). However, in quantum dynamics, nonlinearity is of a restricted kind, and the possibility of obtaining general analytic results providing optimal and robust algorithms for the

feedback control of quantum systems remains an open problem. ■

Further Reading

- Barchielli, A. 1993. Stochastic Differential-Equations and a Posteriori States in Quantum-Mechanics. *Int. J. Theor. Phys.* **32** (12): 2221.
- Brogan, W. L. 1991. *Modern Control Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Carmichael, H. J. 1993. *An Open Systems Approach to Quantum Optics*, Lecture Notes in Physics Monograph 18. New York: Springer-Verlag.
- Doherty, A. C., and K. Jacobs. 1999. Feedback Control of Quantum Systems Using Continuous State Estimation. *Phys. Rev. A* **60** (4): 2700.
- Doherty, A. C., S. Habib, K. Jacobs, H. Mabuchi, and S. M. Tan. 2000. Quantum Feedback Control and Classical Control Theory. *Phys. Rev. A* **62**: 012105.
- Doherty, A. C., K. Jacobs, and G. Jungman. 2001. Information, Disturbance, and Hamiltonian Quantum Feedback Control. *Phys. Rev. A* **63**: 062306.
- Hood, C. J., T. W. Lynn, A. C. Doherty, A. S. Parkins, and H. J. Kimble. 2000. The Atom-Cavity Microscope: Single Atoms Bound in Orbit by Single Photons. *Science* **287**: 1447.
- Mabuchi, H., J. Ye, and H. J. Kimble. 1999. Full Observation of Single-Atom Dynamics in Cavity QED. *Appl. Phys. B* **68** (6): 1095.
- Maxwell, J. C. 1868. On Governors. *Proc. R. Soc. London* **16**: 270.
- Shi, S., A. Woody, and H. Rabitz. 1988. Optimal Control of Selective Vibrational Excitation in Harmonic Linear Chain Molecules. *J. Chem. Phys.* **88**: 6870.
- Wiseman, H. M. 1995. Adaptive Phase Measurements of Optical Modes: Going Beyond the Marginal Q Distribution. *Phys. Rev. Lett.* **75**: 4587.
- Wiseman, H. M., and G. J. Milburn. 1993. Quantum Theory of Optical Feedback via Homodyne Detection. *Phys. Rev. Lett.* **70**: 548.
- . 1994. Squeezing via Feedback. *Phys. Rev. A* **49**: 1350.
- Zhou, K., J. Doyle, and K. Glover. 1996. *Robust and Optimal Control*. Englewood Cliffs, NJ: Prentice-Hall.

Hideo Mabuchi received an A.B. in physics from Princeton University and a Ph.D. from Caltech.



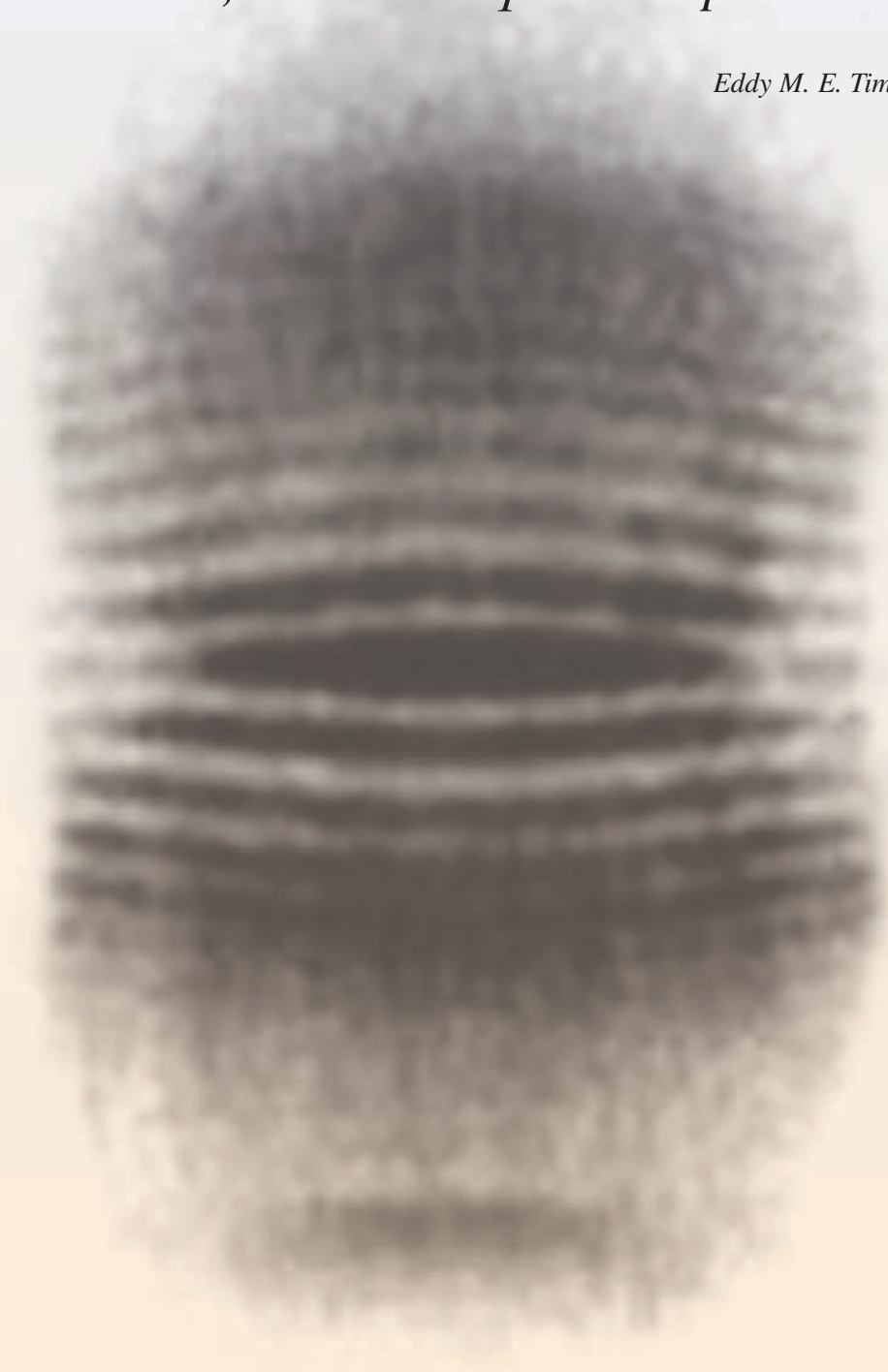
Upon graduating, he became Assistant Professor of Physics at Caltech but spent his first year on leave at Princeton University as a visiting fellow in chemistry. Since returning to Caltech, he has been named an A. P. Sloan Research Fellow, Office of Naval Research Young Investigator, and a MacArthur Fellow. His current research focuses on quantum measurement, quantum feedback control, control-theoretic approaches to the theory of multiscale phenomena, and optical measurement techniques for molecular biophysics.

For biographies of Salman Habib and Kurt Jacobs, see page 125.

Atom-Trap BECs

A new laboratory for studying superfluidity, quantum fluctuations, and other quantum phenomena

Eddy M. E. Timmermans



In October 2001, the field of ultracold-atom physics was honored with the Nobel Prize in physics. It was awarded to Carl Wieman, Eric Cornell, and Wolfgang Ketterle for the creation and study of dilute-gas Bose-Einstein condensates (BECs). Never before had the BEC phase transition, predicted by Einstein more than 70 years earlier, been observed in such a clear and unambiguous realization. By confining neutral atoms in a tiny magnetic trap and cooling them to temperatures only nanokelvins above absolute zero, the Nobel laureates and their colleagues had slowed the atoms down to the point at which the individual wave functions begin to overlap and many thousands of atoms suddenly occupy exactly the same single-particle quantum state. Coaxing bosonic atoms (atoms with integer spin) to condense into this coherent quantum state had been the “holy grail” of the cold-atom physics community for almost two decades. The quest had led to the development of extraordinarily clever trapping and cooling techniques, including Zeeman slowing, magneto-optical trapping, evaporative cooling, and time-orbital potential trapping. The achievement of the first atomic BECs in the summer of 1995 has led to a remarkable sequence of advances that continues unabated.

At first, this article first provides a historical perspective on atom-trap BECs and then focuses on the exciting experiments that are driving the field of cold-atom physics. Our historical overview stresses the long-range coherent properties of BECs and the role BEC physics has played in the explication of superfluidity in liquid helium. In discussing current work, we have selected a line of research and a series of experiments that illustrate the enormous flexibility of the new atom-trap BEC technologies. These experiments were carried out at the Massachusetts Institute of Technology (MIT), Yale University, and Max Planck Institute of Physics in Munich, Germany. Their achievements suggest intriguing prospects for future work in ultracold atomic physics in general and at Los Alamos in particular. In fact, several Los Alamos scientists have already contributed to the development of this field on an individual basis, and we briefly mention those in the concluding section.

The opening figure, produced by Ketterle’s group at MIT, is taken from the paper (Andrews et al. 1997) that provides the starting point for our discussion of the new avenues introduced by these advances. The figure is a direct optical image of two ballistically expanding BECs showing a spatial interference pattern on a macroscopic scale. This pattern is a stunning confirmation that the phase coherence in atom-trap BECs is as complete as in optical lasers, and therefore these condensates can be manipulated and used as atomic lasers, that is, as coherent sources of atomic-matter waves. This is a unique prospect for phase-coherent matter.

After we introduce and resolve an intriguing puzzle regarding the origin of the interference pattern, we turn to a BEC experiment by the group of Mark Kasevich at Yale. This experiment is interesting from a theoretical point of view because the BECs display both laserlike and superfluid aspects of long-range phase coherence. The former is usually reserved for a nonequilibrium system of noninteracting photons, whereas the latter is usually reserved for an equilibrium or near-equilibrium fluid of strongly interacting helium atoms. Specifically, adjacent weakly linked BECs display laserlike spatial interference in a manner that implies Josephson-junction-like phase dynamics between the BECs (Orzel et al. 2001).

The purpose of the Yale experiment was not to probe coherent behavior but to induce and observe quantum fluctuations in the conjugate variables of long-range phase versus localized atom number. The group loaded the BECs into an optical lattice in which the potential barriers separating the lattice wells serve as junctions. By gradually freezing out the motion of the bosons through the junctions and observing the subsequent loss of phase coherence, the scientists were able to infer an increased certainty in the number of atoms in each well, that is, the formation of number-squeezed states. A few months later, the group of Theodore Hänsch in Munich, Germany (Greiner et al. 2002), conducted a beautiful experiment that took this process to its limit. They observed the sudden disappearance of all phase coherence in a BEC trapped in an optical-lattice potential, a direct demonstration of the Mott-insulator phase transition in which a partly coherent state becomes an all-localized state and the tunneling between wells completely stops. This transition is somewhat analogous to the well-known Mott transition from a conducting phase to an insulating phase of electrons in a crystal lattice.

The success of these experiments is due in part to the fact that dilute-gas BECs, with their long coherence lengths and slow evolution times, are readily manipulated and observed with high-precision atomic and optical technologies. Atom-trap BECs have become a remarkably flexible and transparent system for exploring complex many-body phenomena.

In introducing a theoretical view of these developments, we use a “pedestrian” approach to the condensate description, drawing the comparison to single-particle quantum mechanics wherever possible. This approach will make some of the more subtle points of many-body condensate physics accessible to the nonspecialist. We end with an assessment of the atom-trap BEC system for investigating fundamental issues in many-body physics.

Atom-Trap BECs—A Realization of Einstein’s Condensate

Einstein was the first to understand the quantum concept of particle indistinguishability and to realize some of its far-reaching implications. He made the following prediction: When a gas of noninteracting bosons, or particles with integer spin, is cooled below a critical temperature, a significant fraction of the particles will suddenly find themselves in the same lowest-energy single-particle state. (This is an example of a many-body system that is “quantum degenerate,” a term signifying that the system’s behavior is dominated by quantum statistics—that is, the statistics of indistinguishable particles, either Bose statistics for particles with integer spin or Fermi statistics for particles with half-integer spin—as opposed to the Boltzmann statistics of classical systems.) In the limit of zero temperature, all the noninteracting bosons would occupy exactly that same ground state yielding a many-body state that we now call a BEC.

Similarly, in the ground state of a dilute gas of bosons, almost all particles find themselves in the same single-particle quantum state. Much attention has been devoted over the years to the study of such dilute-gas BECs because they are believed to provide a

model for studying superfluidity in a more direct way. The term “superfluidity” denotes a host of low-temperature fluid phenomena such as inviscid, or dissipationless, flow and quantized vortices, all of which contradict our intuition for classical fluid behavior. Interestingly, all condensed-matter superfluids such as helium-4, its fermion cousin helium-3, and the superconductors consist of strongly interacting particles and do not resemble dilute-gas BECs in most of their particulars. However, we believe that their superfluid nature arises from the property of long-range phase coherence, which they share with the dilute-gas BECs. The concept of long-range phase coherence will be discussed later. For now, simply stated, it implies the existence of a complex-valued, single-particle-like wave that characterizes the entire many-body system.

In the case of a dilute BEC, the single-particle-like quantum wave (a wave function that depends on the position of a single particle) can be identified with the wave function of the single-particle state that is occupied, on average, by more than one boson and is also known as the multiply occupied single-particle state.¹ Because almost all particles occupy that single-particle state at zero temperature, the dilute BEC exhibits almost complete coherence. The dilute BEC is then the simplest superfluid system. In contrast, the precise description of the quantum wave coherence of a strongly interacting superfluid is not straightforward. Although it is tempting, for instance, to associate the fraction of the fluid that is superfluid (and can flow without dissipation) with the fraction of the atoms that occupy the lowest-energy single-particle state, that assumption turns out to be wrong. At zero temperature, the helium-3 fluid is all superfluid, whereas only 10 percent of the atoms occupy the zero-momentum state.

Questions regarding the strong interaction effects and the role of quantum fluctuations in reducing the phase coherence and superfluid fraction remain of interest. Against this backdrop, it may be worth noting that the optical-lattice BEC experiments described below give unprecedented control of such quantum fluctuations.

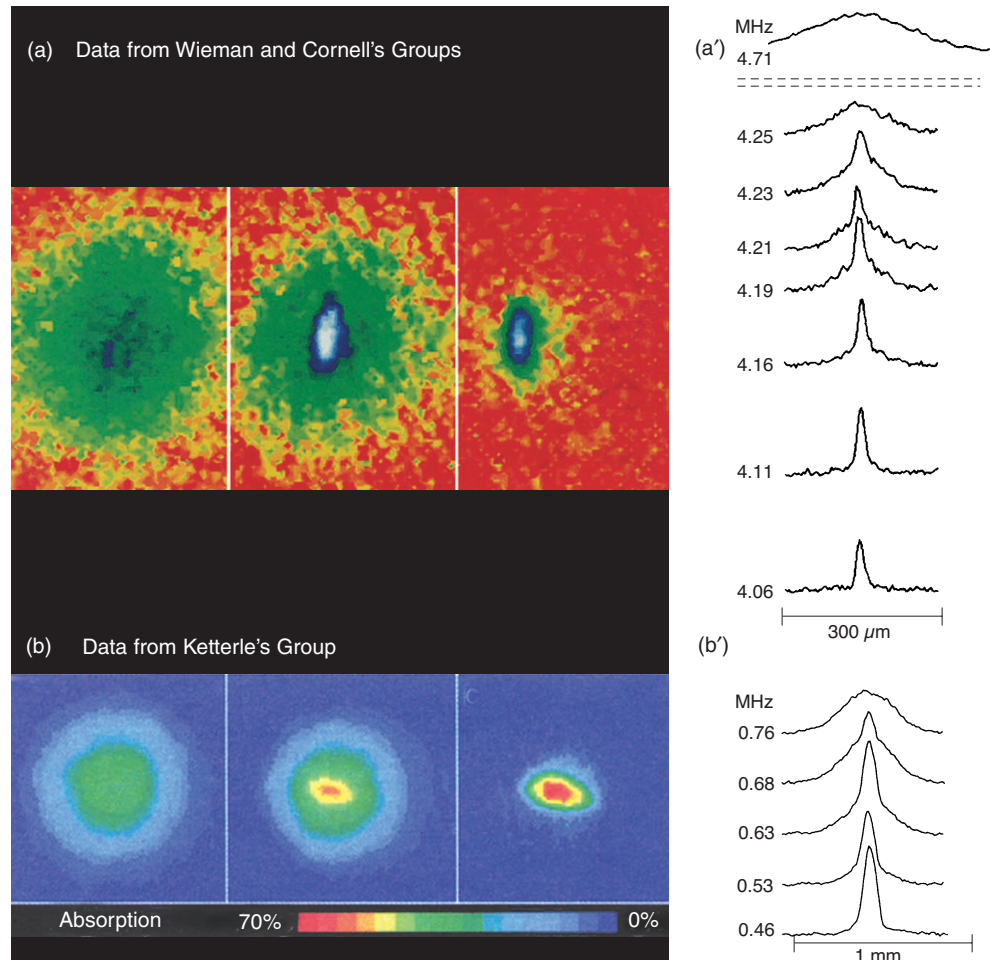
The current atom-trap BECs are dilute in a sense that we will specify shortly. Their experimental achievement represented the first unambiguous realization of dilute BECs. They are made from neutral alkali atoms (sodium, rubidium, lithium, and more recently, hydrogen) that are trapped and cooled with a combination of optical and magnetic fields. (See “Experiments on Cold Trapped Atoms” on page 168 for a description of trapping and cooling processes.) The alkali atoms chosen consist of an even number of fermions (protons, neutrons, and electrons) giving a total spin that has an integer value. These “composite” bosons exhibit the same type of “gregarious” behavior that Einstein predicted for noncomposite bosons. Indeed, the experimenters knew that a BEC had formed when they saw evidence for a sudden increase in the number of atoms occupying the same single-particle ground state at the center of the trap (see Figure 1). This “condensation” is quite different from the familiar liquid-vapor phase transition seen in water, for example. The particle wave functions overlap perfectly, and the behavior of this degenerate Bose-Einstein gas, or condensate, becomes exquisitely sensitive to the interparticle interactions even if the system is dilute. The spatial extent of the multiply occupied single-particle wave function is determined by the competition of the effective interparticle repulsion and the trapping potential that confines the atoms. In present-day experiments, the size of the BEC can be as large as one-tenth of a millimeter. In other words, the multiply occupied single-atom wave function describing the BEC is macroscopic.

Although Bose-Einstein condensation had never been directly observed before 1995, this phase transition served as a textbook example in statistical mechanics (Huang 1987) because it is one of the few phase transitions that can be described analytically. As Einstein himself stressed (Pais 1979), this remarkable transition follows solely from the quantum-mechanical concept of particle indistinguishability, unlike the usual phase

¹ A multiply occupied single-particle state is mathematically expressed in Equation (1) of this article.

Figure 1. The First Atom-Trap BECs

Some of the first signatures of Bose-Einstein condensation were obtained in a dilute gas of trapped rubidium atoms in the groups of Wieman and Cornell. Shown in (a) are the shadow (absorption) images of the density profile of the trapped atoms and in (a') the cross sections of the local density. Both data sequences were obtained with varying values of the cutoff energy used in the evaporative cooling, the final stage in cooling the trapped atoms. In evaporative cooling, atoms of energy above the cutoff, indicated in megahertz, were removed from the trap. As the cutoff energy decreases, the final temperature to which the system equilibrates is lowered. Below a critical value, a sharp peak appears in the density profile, a signal that Bose-Einstein condensation has occurred. As the gas was contained in an asymmetric (cigar-shaped) trap, the shape observed in (a) provides an independent signature. The left-most frame shows a spherically symmetric thermal cloud; the middle frame shows an asymmetric density spike corresponding to the condensate surrounded by a thermal cloud; and the rightmost frame shows the final density spike in which most of the atoms have Bose-condensed. (b) These shadow images from Ketterle's group show a BEC in sodium. The number of trapped atoms is greater than that in (a) by about a factor of 100. The density of the condensate grows with decreasing temperature from left to right. (b') These density plots show cuts through an atomic cloud as the condensate develops. Note that the spatial extent of the condensate is about 0.1 mm. The size reflects the macroscopic nature of the system. It increases with the scattering length defined in the text.



[Figures 1(a) and 1(a') are reprinted with permission from Anderson et al. *Science* 269, page 199. Copyright 1995 American Association for the Advancement of Science. Figures 1(b) and 1(b') are reproduced with permission from the American Physical Society.]

transitions, which result from a competition between interactions and entropy (disorder).

The neutral atom-trap systems are extremely dilute. Like billiard balls, they feel each other's presence only when they are separated by a distance equal to or less than a particular length. This length, known as the scattering length a , takes on different values for different atomic species—or even for the same species in different atomic states—but for most of the trapped neutral alkali atoms, its value is positive (reflecting an effectively repulsive force between the particles), and it tends to be about 1 nanometer. We characterize the “diluteness” of the gas by visualizing the atoms as hard spheres of radius a and computing the fraction of the total volume occupied by the spheres, $(4\pi/3)na^3$, also called the “packing fraction.” In the current atom-trap BECs, the packing fraction ranges from one part in a million to one part in a billion.

At that diluteness, almost all atoms are phase coherent in the zero-temperature Bose-condensed state, somewhat in the manner that the photons produced through stimulated emission into a single mode of an optical-laser cavity are phase coherent. That is, all particles behave according to the same coherent wave function, and the particles can exhibit macroscopic interference. Contrary to the optical-laser system, the BECs consist of mutually interacting particles that are conserved (that is, the total number of atoms remains constant) and that can relax to an equilibrium state, in which case the long-range phase coherence gives rise to superfluid behavior. Indeed, in the last three years,

experiments have definitively shown that the atom-trap BECs exhibit the defining behavior of a superfluid such as sustained superflow (or dissipationless flow), zero resistance to an object moving through the condensate, and quantized vortices.

Most BEC experiments are carried out with no more than a hundred thousand to a few million atoms. The difficulties encountered in increasing the particle number currently limits the prospects for practical applications somewhat. On the other hand, the atom-trap BEC technology has become fairly routine—more than 20 experimental groups have achieved BECs by now. The extraordinary flexibilities offered by the available atomic, molecular, and optical technologies, as well as by the imaging techniques, provide the BECs with advantages that are unique in low-temperature physics.

Aspects of BEC Dynamics

We will explore a bit further the two quantum concepts that are central in understanding BECs and the sense in which superfluid behavior of the BECs represents the behavior seen or inferred in liquid helium and other systems, including nuclei, subnuclear systems produced in accelerators, and neutron stars. Those two central concepts are particle indistinguishability and coherent wave behavior.

Particle Indistinguishability. It was Einstein who realized that the statistics Bose devised to understand the Planck spectrum of black-body radiation involved counting the number of ways in which particles (in that case, photons) can be distributed over single-particle states (called “subcells” in Einstein’s thermodynamic treatment). The Bose counting presumed the particles to have a distinctly nonclassical quality. Whereas the trajectories of classical particles can always be followed so that the particles can be distinguished from each other, Bose counting assumed particles to be fundamentally indistinguishable. Einstein extended the counting technique for photons, whose particle number is not conserved, to a gas of conserved noninteracting particles, and he showed that the indistinguishability implies a sudden increase in the number of particles occupying the specific subcell/single-particle state of lowest energy: the BEC phase transition.

Coherent Wave Behavior. A BEC’s coherent wave behavior follows directly from the time evolution of the multiply occupied single-particle state. In quantum mechanics, the one-particle system evolves according to Schrödinger’s wave equation. As a consequence, the single-particle system can exhibit the type of interference seen in Young’s classic double-slit experiment, which proved that light was a wave phenomenon (see the box “The Double-Slit Experiment”). In the quantum interpretation, light and atoms exhibit both particle and wave behavior, and the interference results from the uncertainty in knowing which of two possible trajectories the particle or the photon followed in reaching the detector. (Put another way, the particle can simultaneously follow two different paths to reach the screen; that is, it can exist in a superposition of probability amplitudes A_1 and A_2 , one for each path. The probability of finding the particle at the detector is given by the square of the amplitude $|A_1 + A_2|^2$, which exhibits interference that is due to the $A_1 A_2^* + A_1^* A_2$ contribution.) Depending on the location at which the particles hit the detector, the probability amplitudes for each path add up constructively or destructively, respectively increasing or decreasing the probability.

As explained in the box, the observation of an interference pattern, even with light, can represent an experimental challenge. Many particles (or photons) must pass through the slits for the pattern to be seen, and if the particles (photons) occupy different single-particle states, the interference washes out, and the probability becomes a single blob without the spatial oscillations that signal interference. In the BEC case, as in an optical-

The Double-Slit Experiment—A Quantitative Measure of Coherence

In 1802, Young devised and performed the double-slit experiment, which disproved Newton's particle theory of light and established unequivocally that light is a wave phenomenon. In that experiment, two holes punched in a screen allowed incident light to pass through. The light intensity reaching a second screen located behind the first was then recorded, and under the right conditions, it was possible to observe interference fringes (an intensity pattern that oscillates in space), giving unmistakable proof of the wave nature of light.

To understand the origin of the interference fringes, we imagine the light to be perfectly monochromatic (characterized by a single wavelength or frequency) and to be emitted in a direction perpendicular to the screens from a point source an infinite distance away (see Figure A). In that case, the incident light consists of plane waves with wave fronts parallel to the screen. The light reaching a specific position on the second screen has traveled in a straight line from either hole, and the difference in distance traveled determines the difference in phase of both light rays reaching the screen. If the difference in distance traveled by each ray is equal to an integer number of wavelengths, the waves originating from each hole are in phase, which means that their instantaneous electric-field vectors point in the same direction. The total electric field, which is the vector sum of both fields, then has a magnitude equal to the sum of the magnitudes. In contrast, if the difference in distance is equal to an odd number of half-wavelengths, the waves are out of phase, meaning that the electric-field vectors of the rays that passed through the different slits point in opposite directions and that the magnitude of their vector sum is less than that of the light from a single hole. In fact, they can completely cancel each other out, giving a vanishing intensity. In the first case, the waves are said to add up constructively, and the intensity, which is proportional to the square of the magnitude of the total electric-field vector, appears bright; in the latter case, the waves add up destructively, and the intensity appears dim. Varying the position on the second screen causes the difference in distance from both holes to vary and the intensity to go through a series of maxima and minima, corresponding to, respectively, constructive and destructive interference.

In a realistic two-slit experiment, the incident waves are not perfectly monochromatic, and the source of light is not a perfect point source. Whether the interference pattern can be distinguished in the recorded intensity actually depends on the details of the experiment, such as the distance between the slits. Loosely speaking, optical coherence refers to the ability of the light to exhibit such interference. Mathematically, the contrast is specified by

measurements of the highest (I_{\max}) and lowest (I_{\min}) intensities. The visibility of the fringes, defined as the ratio $(I_{\max} - I_{\min}) / (I_{\max} + I_{\min})$, provides a measure of light coherence. For laser light, the slits can be as far apart as the width of the laser beam and still produce an interference pattern with a visibility near unity. In the quantum description of the laser, nearly all photons are said to be in the same state. In contrast, thermal light contains photons in different states, each of which would give a different interference pattern with interference fringes at different positions. The recorded pattern is a sum of all the interference patterns, and the fringes at different positions can wash each other out.

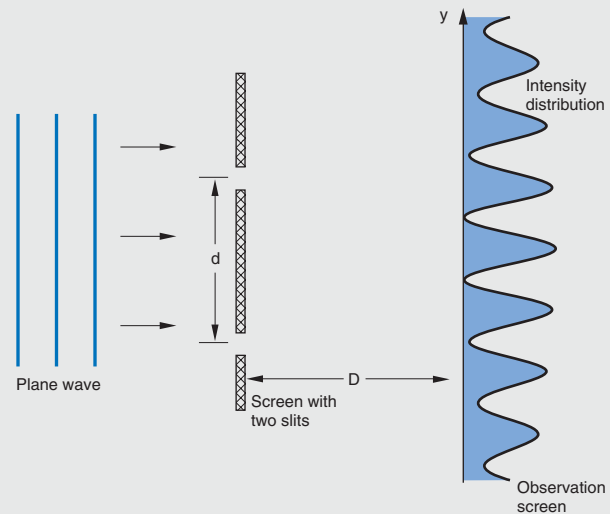


Figure A. Diagram of Double-Slit Experiment

A plane wave incident on the first screen passes through the two slits and is stopped by the second screen. The light intensity at a specific position on the second screen depends on the difference in the path lengths traveled by the light waves emanating from the two slits to that position. If the path length difference is equal to an odd number of half wavelengths, the spot appears dim (low intensity); if it is equal to an integer number of wavelengths, the spot appears bright (high intensity). The path length difference varies along the straight line shown in the plane of the second screen. Along this line, the intensity passes through positions of constructive and destructive interference, giving an oscillatory intensity variation, called interference fringes.

laser system, most particles occupy the same state so that the many-particle system exhibits the interference pattern of the single-particle system. We call this property “coherent wave behavior.” As mentioned previously, it is the essential property that the weakly interacting BEC has in common with the strongly interacting superfluids such as helium.

Classical or Mean-Field Description of BEC Dynamics. Current atom-trap BECs have packing fractions of about one part in a million to one part in a billion. At that diluteness, almost all the neutral atoms of a near-equilibrium system at near zero temperature occupy the same single-particle state. The many-body system can therefore be approximated by an N -particle wave function consisting of a product of single-particle wave functions:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; t) \approx \chi(\mathbf{r}_1; t) \chi(\mathbf{r}_2; t) \dots \chi(\mathbf{r}_N; t), \quad (1)$$

where the single-particle χ -function is a complex-valued quantity:

$$\chi(\mathbf{r}; t) = |\chi(\mathbf{r}; t)| \exp(i\theta(\mathbf{r}; t)). \quad (2)$$

In 1927, shortly after the discovery of quantum mechanics, Erwin Madelung pointed out that the behavior of the single-particle wave function was analogous to that of a fluid in which $|\chi(\mathbf{r}; t)|^2$ plays the role of the single-particle density and $(\hbar/m)\nabla\theta$ is associated with a velocity. Similarly, in BEC physics, where the single-particle wave function is multiply occupied, the phase of the single-particle wave function, θ , plays a crucial role in the theory as the single phase that gives rise to all the coherent wave phenomena discussed below. In particular, its gradient describes the velocity associated with the dissipationless flow observed in superfluid systems.

The product state in Equation (1) is a special case of the Hartree-Fock Ansatz for the many-body wave function of identical particles, and it evolves according to a Hartree-Fock equation of motion. If the boson particles of mass m experience an external trapping potential V , so that the potential energy of a single boson at position \mathbf{r} is $V(\mathbf{r})$, and if the bosons interact with each other through an interaction potential v , so that a pair of bosons located respectively at \mathbf{r} and \mathbf{r}' experience an additional energy $v(\mathbf{r} - \mathbf{r}')$, then the Hartree-Fock equation takes on the following form:

$$i\hbar \frac{\partial \chi}{\partial t} = \left[-\frac{\hbar^2 \nabla^2}{2m} + V(\mathbf{r}) \right] \chi + [N-1] \int d^3\mathbf{r}' v(\mathbf{r} - \mathbf{r}') |\chi(\mathbf{r}'; t)|^2. \quad (3)$$

Because the interaction between neutral atoms in a BEC has a much shorter range than the length scales on which the atom-trap BECs vary, we can approximate the interparticle potential by an effective contact interaction, $v(\mathbf{r} - \mathbf{r}') \rightarrow \lambda \delta(\mathbf{r} - \mathbf{r}')$, where the interaction strength λ is proportional to the scattering length a : $\lambda = (4\pi\hbar^2/m)a$. In addition, the number of particles is large enough to allow approximating $(N-1)$ by N . We then introduce the condensate field Φ as $\Phi = N^{1/2}\chi$ so that $|\Phi|^2$ represents the

particle density, as it does in the single-particle case. With these quantities, the Hartree equation for atom-trap BECs takes on the form of the celebrated Gross-Pitaevski equation:

$$i\hbar \frac{\partial}{\partial t} \Phi = \left[-\frac{\hbar^2 \nabla^2}{2m} + V + \lambda |\Phi|^2 \right] \Phi . \quad (4)$$

This equation, first derived by Pitaevski to treat superfluid vortices in a full quantum description, has been very popular in many fields of physics (and even biology). In spite of its simplicity, it has solutions that exhibit crucial nonlinear physics phenomena such as solitary waves, self-focusing, and self-trapping. As a result, the atom-trap BECs can also be regarded as new laboratories for studying nonlinear dynamics.

Describing the physics of BECs by means of the Gross-Pitaevski equation—Equation (4)—is known as “making the mean-field approximation” or “working in the classical approximation.” The term “classical” may appear out of place because Equation (4) implies that matter has wavelike behavior, and it implicitly contains the Planck constant. Nevertheless, this equation also follows from the Lagrange equations of the corresponding classical field theory without any quantization condition. The Gross-Pitaevski equation gives a classical description of BECs in the same sense that Maxwell’s equations provide a classical description of photon dynamics. Perhaps most significantly, the Gross-Pitaevski equation provides the simplest possible description of a superfluid system, and the mean-field approximation (which for BECs is equivalent to assuming a product wave-function solution) captures many of the essential features of superfluidity. For instance, the mean-field treatment predicts a dispersion relation, or excitation spectrum, that satisfies Landau’s criterion for dissipationless flow (a criterion to which we refer below). On the other hand, the Gross-Pitaevski equation is certainly not as general as the phenomenon of superfluidity. Although some long-range behavior of the helium superfluids and superconductors can be qualitatively understood when this equation is invoked, the atom-trap BECs are the only systems quantitatively described by it. Moreover, the classical description also breaks down for BECs, for example, when quantum fluctuations become important, as they do in the experiments described at the end of this article. Those experiments involve number-squeezed states and the Mott transition from a coherent, or superfluid, state to a localized state.

The Coherent Wave Nature of Superfluidity

The term “superfluidity” was first applied to a very low temperature phase of liquid helium. In 1938, Peter Kapitza and, independently, John Allan and Donald Misener discovered that below a critical temperature of 2.2 kelvins, liquid helium-4 flows without measurable dissipation through capillary tubes. It seemed that this low-temperature phase of helium-4, called He_{II}, is not governed by the usual laws of classical fluid dynamics. Subsequent experiments uncovered other counterintuitive phenomena in He_{II}, including the fountain effect, perfect heat conductivity, and persistent circular flow. Superfluidity is now the name for both this collection of phenomena and the state of matter responsible for them.

The superfluid state was so unusual and its mechanism so difficult to discern in the relatively inaccessible medium of a strongly interacting fluid that its origin remained a matter of continuing controversy for more than two decades.

Is He_{II} like a BEC? Noting that helium-3, the fermion cousin of helium-4, did not undergo a phase transition to a superfluid at similar temperatures, Fritz London suggested in 1938 that the He_{II} transition is intimately related to the boson nature of the helium-4 atoms. He further proposed that the He_{II} superfluid is, in a generalized sense, a BEC. Of course, being a strongly interacting fluid, the helium system cannot be characterized by the assumption that all atoms occupy the same single-particle state. Nevertheless, London (1938) argued that “some of the general features of the degenerate ideal Bose-Einstein gas remain intact, at least qualitatively, for this liquid.” He also offered support for his thesis by calculating the BEC critical temperature for the helium density, which came out to 3.13 kelvins, remarkably close to the He_{II} transition temperature of 2.12 kelvins, measured in 1933. Although the latter agreement is largely fortuitous, London’s words sound almost prophetic in retrospect: He hinted that the superflow in He_{II} was a macroscopic quantum current brought about by changes in the boundary conditions.

The Two-Fluid Description of He_{II}. Following a different track, Lev Landau and, independently, Laszlo Tisza (who was, in fact, partly motivated by London’s views) proposed the two-fluid model of He_{II}, in which one component is an inviscid, irrotational superfluid that does not carry entropy. This model explained the observed effects and also correctly predicted new superfluid phenomena, such as second sound. Landau used very general assumptions to derive a criterion for superfluidity and an expression for the critical velocity above which dissipation would set in. The critical-velocity calculation, although ultimately incorrect, captured the main features of persistent flow, and a generalized form of the Landau criterion is still of great use in explaining critical velocities for superfluidity. Nicolai Nicolaevich Bogoliubov showed that a weakly interacting BEC satisfies Landau’s criterion for superfluidity, but Landau continually resisted the notion that the superfluid should be associated with a BEC.

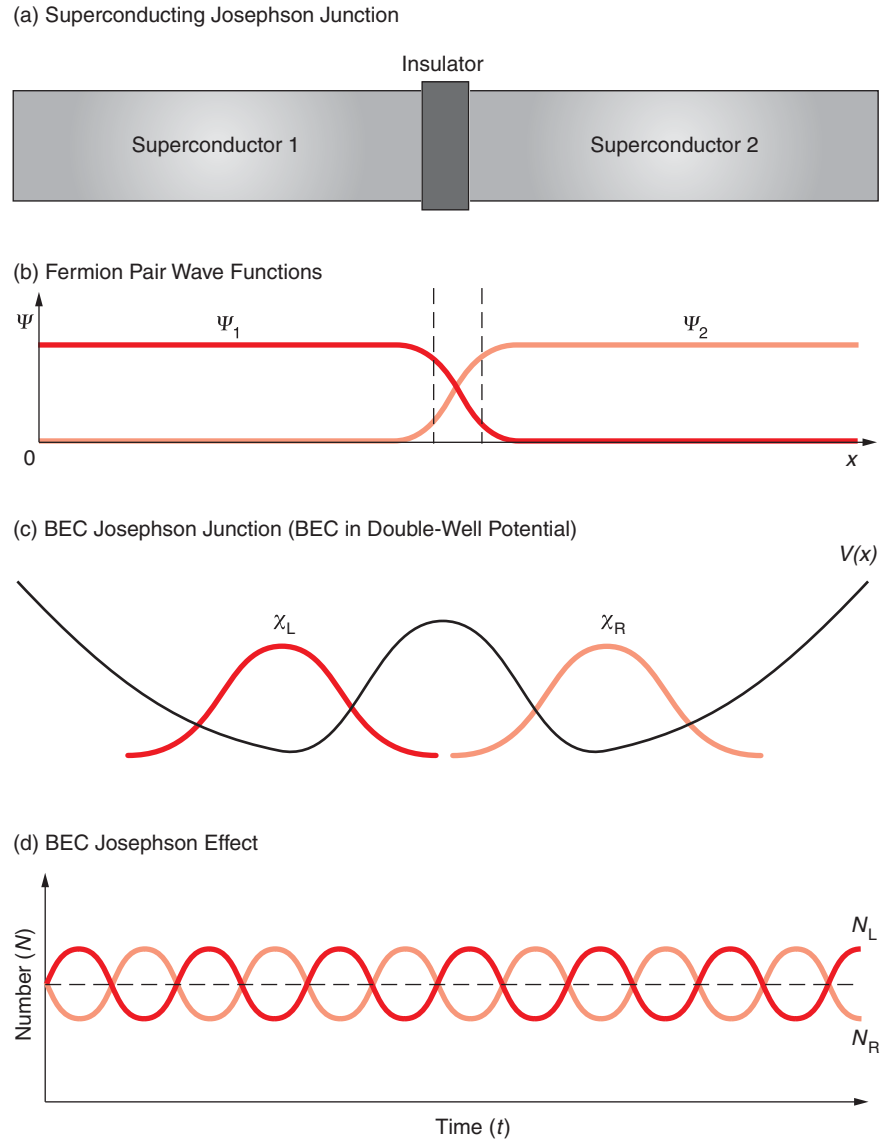
The BEC Description Revisited. Finally, Oliver Penrose (1951) and then Penrose and Lars Onsager (1956) proposed the currently accepted point of view that superfluidity is a macroscopic manifestation of coherent (hence, single-particle-like) quantum-wave behavior. This description does not contradict the two-fluid model but supersedes it in the sense that the coherent quantum-wave behavior includes phenomena, such as quantized vortices and Josephson effects, which find no explanation in the two-fluid model.

As previously mentioned, the single-particle quantum wave behavior, which is compatible with and can be described as fluidlike behavior, had been pointed out by Madelung in 1927. In his pioneering paper of 1951, Penrose derived the equation for the off-diagonal density matrix of the many-body helium fluid and then drew on Madelung’s analysis of the single-particle wave function to associate the long-range part of that off-diagonal density matrix with the superfluid component of the two-fluid model. In essence, Penrose identified quantum wave coherence as the essential feature responsible for both superfluidity and the BEC-like behavior conjectured by London.

As the understanding grew that superfluidity was an outcome of quantum wave coherence, the intimate connection between superfluidity and superconductivity was realized. We now understand both phenomena to be caused by coherent quantum-wave behavior, that is, many identical particles or units whose behavior can be described by the same single-particle wave function. For a superfluid, the single unit that exhibits the quantum wave behavior is a boson particle; for a superconductor, it is a pair of fermions. Much as we regard a superfluid as a BEC of boson particles, we can regard a superconductor as a BEC of fermion pairs. Not surprisingly, therefore, the fields of

Figure 2. Josephson Junctions and the Josephson Effect for BECs

The diagrams show (a) two superconductors separated by a thin barrier and (b) the overlap of the coherent single-particle wave functions that describe each superconductor in the neighborhood of the junction. In 1962, Brian Josephson showed that, under certain conditions, quantum mechanical tunneling of electron pairs could occur through the barrier. If the two wave functions differ by a phase, a direct current of electron pairs will flow through the barrier, or junction. If a voltage is placed across the junction, the phase difference varies periodically in time, causing an alternating current to flow across the junction. (c) A neutral-atom BEC trapped in a double-well potential behaves like a superconducting Josephson junction. The potential barrier created by a laser beam acts like the insulating barrier between the superconductors. (d) The BEC junction is predicted to exhibit the Josephson effect. For instance, a sudden change in the chemical potential of one of the BECs would initiate an oscillation in the number of particles in each well. The frequency of the oscillation is determined by the difference of the chemical potentials.



superfluidity and superconductivity share a number of phenomena that stem directly from their coherent wave nature. Two of these coherent phenomena, Josephson junctions and quantized vortices, have recently been studied in atom-trap BECs and are briefly described next.

Josephson Junctions. In the 1960s, the physics of superconducting Josephson junctions provided evidence for the coherent wave nature of superconductors. The Josephson junction is a weak link, such as a thin insulator, connecting two indistinguishable superfluids or superconductors—see Figures 2(a) and 2(b). One manifestation of the Josephson “effect” is an alternating current flowing through the weak link when both sides of the junction are kept at different chemical potentials by, for instance, the introduction of a potential difference over the junction.

In an ordinary electronic circuit, the potential difference sets up a direct current (dc), which flows from the region of high chemical potential to that of low chemical potential. In contrast, in a coherent-wave superfluid system, the rate for bosons or fermion

pairs to tunnel through the potential barrier of the junction depends sinusoidally on the phase difference between the single-particle-like wave function on either side of the junction. That phase difference increases linearly with time in the presence of a potential difference, giving an alternating current that oscillates at the frequency corresponding to the chemical potential difference.

In the original condensed-matter Josephson junctions, the superfluids were superconductors. In such cases, the bosons tunneling through the junction are electron pairs, and the current is a charge current, which is easily and accurately measured. In helium superfluids, on the other hand, the weak link is difficult to make, and the observation of a weak neutral current presents a nontrivial experimental challenge, which was only recently met (Packard 1998).

The direct analogue of the Josephson junction in atom traps is an atomic BEC trapped in a double-well potential—see Figures 2(c) and 2(d). The challenge of observing the Josephson effect in this system, however, is similar to the problem encountered in observing Josephson oscillations in helium superfluids: How can one measure small-amplitude oscillations of neutral-particle populations? In the last section, we show how atom-trap BEC technology made possible a unique solution to the problem of observing Josephson phase dynamics.

Quantized Vortices. Quantized vortices are another coherent wave phenomenon unique to superfluids and superconductors. In classical fluids, vortices are long-lived flow patterns in which the particles whirl around an axis, all with the same angular momentum. In a superfluid, a superflow that similarly whirls around an axis can be set up by a characteristic variation of the coherent wave function: the phase of the wave function varies cylindrically around the vortex axis. For the wave function to be single-valued, it must return to its initial value after a full rotation around the axis; that is, its phase must have changed by 2π or by $2\pi n$, where n represents an integer number. This constraint implies that the angular momentum of superfluid vortices is quantized with allowed values equal to $n\hbar$ —see Figure 3(a).

Quantized vortices in helium were observed by William Vinen and by George Rayfield and Frederick Reif, and their observations provided further support for the coherent wave behavior of the helium superfluid. In atom-trap BECs, the long-lived metastable vortex structures were created and studied in laboratories at the Joint Institute for Laboratory Astrophysics (JILA) at Boulder, Colorado, in the groups of Wieman and Cornell; at the École Normale Supérieure in Paris, in the group of Jean Dalibard; at MIT in the group of Ketterle; and at Oxford University, England, in the group of Chris Foot—see Figure 3(b). A direct measurement of the angular momentum of the vortices, by Dalibard's group, experimentally confirmed the quantization of BEC vortices. In addition, at MIT, rapid advances in BEC technology led to the creation of vortex lattices (also called Abrikosov lattices) in atom-trap BECs with up to 160 vortices and to the detailed observation at both MIT and JILA of the intricate dynamics of vortex formation and decay.

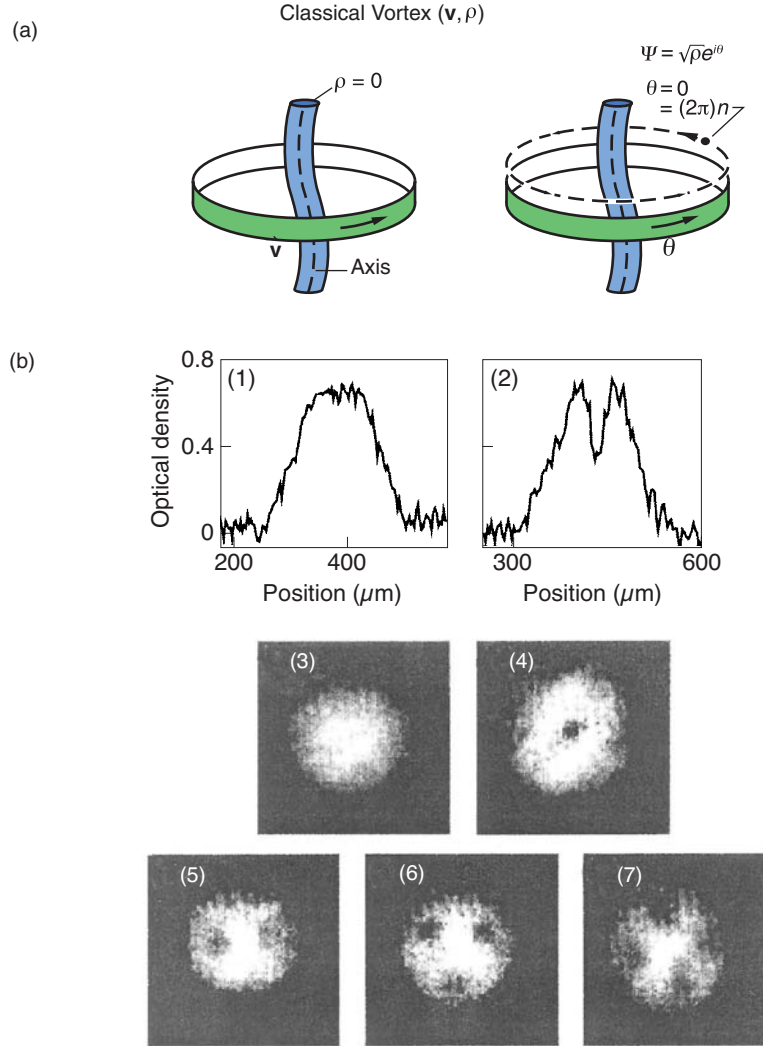
BEC Interference—A Demonstration of Wave Coherence

In optical systems, long-range phase coherence is easily demonstrated through the double-slit experiment. In fact, the sharpness of the interference fringes produced in that experiment is used as the standard measure of optical coherence. In contrast, condensed-matter systems give mostly indirect signatures of wave coherence—quantized vortices and Josephson effects—although observations and applications of temporal interference in superconductors do exist (for example, in superconducting quantum interference devices, or SQUIDS).

Figure 3. Quantized Vortices

(a) In a superfluid, the phase of the wave function for a vortex must increase by 2π on each revolution, which implies that the angular momentum of the vortex must be an integer multiple of \hbar , or $n\hbar$. (b) Several experimental groups have created and imaged quantized vortices in atom traps. The transverse absorption images (Madison et al. 2000) are of a condensate of about 10^5 rubidium-87 atoms at a temperature below 80 nK. This condensate has been stirred with a laser beam at various rotational frequencies. Above a critical rotational frequency, vortex filaments appear. Plots 1 and 2 show the variation in optical thickness along the horizontal axes of the clouds imaged in plots 3 and 4, respectively. The cloud stirred at 145 Hz (shown in plot 3) contains no vortex filament, whereas the cloud stirred at 152 Hz (shown in plot 4) contains one vortex filament. In plots 5, 6, and 7, the condensate was stirred at rotational frequencies of 169, 163, and 168 Hz, respectively.

(Reproduced with permission from *The American Physical Society*.)

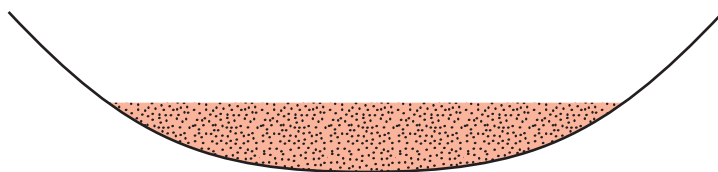


Thus, when Ketterle’s group at MIT observed the spectacular interference pattern shown in the opening illustration, they brought an unusual message: BECs are superfluids that can manifest their long-range phase coherence in an optical-laser-like manner of spatial interference. Michael Andrews and collaborators later (1997) argued that the interfering BEC experiment demonstrated the first atom laser (albeit in a form that, as of yet, is not necessarily useful to applications). Their demonstration suggests that the simultaneous appearance of superfluid and laserlike aspects of long-range phase coherence might one day yield particularly potent applications of BECs.

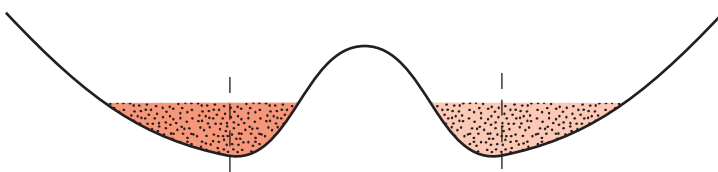
The MIT Experiment. Figure 4 outlines the experimental procedure used by the MIT group. First, an off-resonant laser beam is passed through the center of an atom trap, which effectively creates a double-well potential. The atoms are then cooled and Bose-condensed into two BECs, one on either side of the potential barrier—see Figure 5(a). Because the height of the barrier significantly exceeds the chemical potential of either BEC, the two BECs are independent.

When the trapping potential was switched off, the two BECs expanded freely and started overlapping spatially. Using two laser pulses in succession, the MIT group imaged the local density of atoms in a 100-micrometer-thick slice within the region of

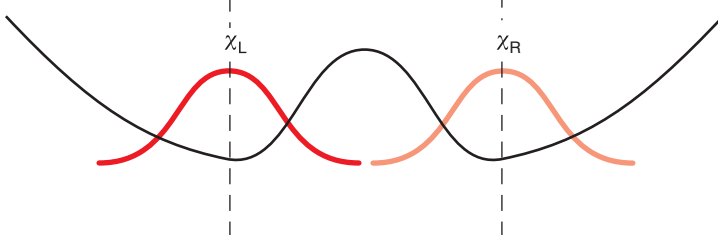
Stage 1: Sodium atoms are contained in a single-well trapping potential.



Stage 2: A laser beam repels the atoms and creates a trapping potential with a double-well shape.



Stage 3: The atoms are cooled below the critical temperature of the phase transition to BECs.



Stage 4: The trapping potential is suddenly removed, and the BECs expand and overlap.

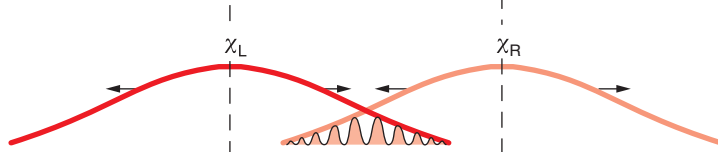


Figure 4. Procedure for Creating BEC Interference
In the BEC interference experiment conducted at MIT (Andrews et al. 1997), sodium atoms were contained in a cigar-shaped trap (stage 1). In the second stage, a laser beam focused on the center of the initial trap repelled the atoms from that region, creating an overall atomic potential that has a double-well shape. In the third stage, the atoms were cooled below the critical temperature T_C of the BEC phase transition. The height of the potential barrier separating the wells greatly exceeded the thermal energy $k_B T_C$ (where k_B denotes the Boltzmann constant) and the chemical potentials of the BECs that are formed in the left (L) and right (R) wells. (The wave functions for the two BECs are labeled χ_L and χ_R .) Under these conditions, the two BECs are independent of each other in the sense that they cannot “know” each other’s phase. When the trapping potential is suddenly removed in stage 4, both BECs expand and then overlap. Images of the atomic density of the overlapping BECs show macroscopic interference fringes of high visibility.

overlap. The first laser pulse pumped the BEC atoms in the selected slice from state $|1\rangle$ to a different hyperfine state $|2\rangle$. The second laser, tuned near a resonant transition from state $|2\rangle$ to state $|3\rangle$ and pointing more or less perpendicular to the plane of the slice, imaged the density of atoms in state $|2\rangle$. The image showed a highly visible, regular pattern of clearly separated interference fringes of macroscopic size (40 micrometers)—see Figure 5(b). The visibility of the fringes (defined in the box “The Double-Slit Experiment”) ranged from 20 to 40 percent. By characterizing their optics, the experimentalists inferred that the actual visibility of the density fringes ranged from 50 to 100 percent. The density fringes are defined as $(\rho_{\max} - \rho_{\min})/(\rho_{\max} + \rho_{\min})$, where ρ_{\max} and ρ_{\min} denote the maximum and minimum densities if observed with an ideal imaging technique. The high visibility of the observed fringes indicates that the entire many-body system behaves as a coherent wave.

What Produces the Interference Fringes? Unquestionably (by definition, in fact), macroscopic interference fringes indicate coherence in the usual optical sense. But how the observed interference fringes relate to the coherence of the expanding BECs is a matter of considerable subtlety, as will be explained. Under the experimental conditions of independent BECs, the single-particle density matrix, as we show below, does not

Figure 5. Sodium Atom BECs and Their Interference

(a) Phase contrast images of a single Bose condensate (upper panel) and double Bose condensates were taken in the magnetic trap of the MIT group. An argon ion laser that was focused into the center of the trap created a double-well potential. Changes from 7 to 43 mW in the power of the laser-light sheet caused the distance between the two condensates to vary. (b) The interference pattern of two expanding condensates was observed after a 40-ms time of flight for two different powers of the argon-laser-light sheet (raw-data images). The periods of the fringes were 20 and 15 μm ; the laser powers were 3 and 5 mW; and the maximum absorptions were 90% and 50%, respectively, for the left and right images. The fields of view were 1.1 mm horizontally by 0.5 mm vertically. The horizontal widths were compressed fourfold, a condition that enhances the effect of the fringe curvature. For the determination of the fringe spacing, the dark central fringe on the left was excluded.

(Reprinted with permission from Andrews et al. *Science* 275, pages 638 and 639. Copyright 1997 American Association for the Advancement of Science.)

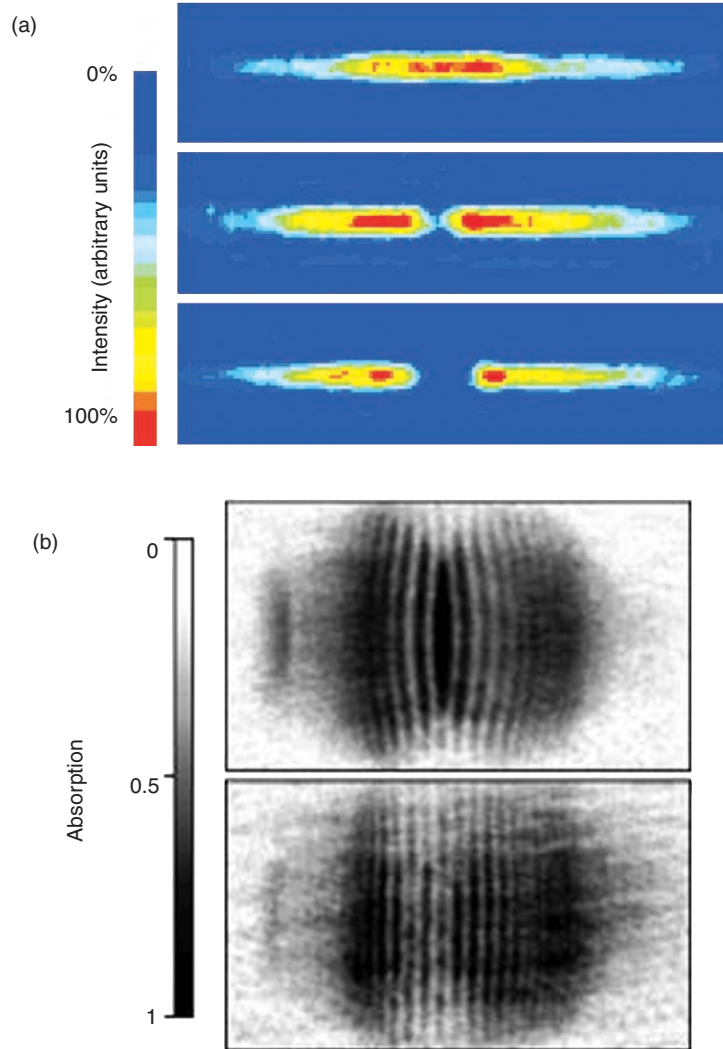


exhibit interference. Why then does the recorded image show fringes? The resolution, as we show for a special case, depends on the fact that the image does not record the single-particle density.

The Case of BECs with Definite Particle Number. As reported by Andrews et al. (1997), the potential barrier separating the two BECs was five times higher than the energy corresponding to the critical temperature for the BEC phase transition and 50 times higher than the chemical potentials of the BECs in each well. Under those conditions, the state of the double-well BEC system is indistinguishable from that of two BECs that were condensed in separate traps at an infinite distance from each other and then brought together. In principle, we can therefore know exactly how many particles occupy each of the two BECs. That is, the system is in a number state. The single-particle density of this double-well number state $\rho_{1(N)}$ does not exhibit interference, a point we now demonstrate for a simplified double-well number state with only two particles.

We call the single-atom state centered in the right well $\chi_R(\mathbf{r})$ and the single-atom state centered in the left well $\chi_L(\mathbf{r})$, where \mathbf{r} denotes the center-of-mass position of the trapped atom. Thus, a two-particle number state with one atom in each well is represented by a wave function $\Psi_{(N)}$:

$$\Psi_{(N)}(\mathbf{r}_1, \mathbf{r}_2; t) = 2^{-1/2} [\chi_L(\mathbf{r}_1; t) \chi_R(\mathbf{r}_2; t) + \chi_R(\mathbf{r}_1; t) \chi_L(\mathbf{r}_2; t)] . \quad (5)$$

When the external potential is switched off, the two-particle wave function, to a close approximation, remains of the form in Equation (5), with χ_L and χ_R evolving as freely expanding single-particle wave functions that are mutually orthogonal. The corresponding single-particle density $\rho_{1(N)}$ at a given time t ,

$$\begin{aligned} \rho_{1(N)}(\mathbf{r}; t) &= \int d^3 r_2 \left| \Psi_{(N)}(\mathbf{r}, \mathbf{r}_2; t) \right|^2 \\ &= \frac{1}{2} \left[\left| \chi_L(\mathbf{r}; t) \right|^2 + \left| \chi_R(\mathbf{r}; t) \right|^2 \right], \end{aligned} \quad (6)$$

is equal to an incoherent average of the densities of the individual expanding single-particle wave functions. Generally, the single-particle densities expand smoothly—a free-particle Gaussian wave function (for instance, if the χ -wave-functions start out as ground-state functions of harmonic oscillator potentials) remains Gaussian—so that $\rho_{1(N)}(\mathbf{r}; t)$ does not exhibit spatial oscillations.

The Case of a Mutually Coherent State of the Double-Well System. In contrast, had a single-well system containing both particles in its center-of-mass ground state been split adiabatically, the resulting double-well system would be in a mutually coherent state. This particular mutually coherent state would be a product of single-particle wave functions of the type $2^{-1/2}[\chi_L(\mathbf{r}; t) + \exp(i\alpha) \chi_R(\mathbf{r}; t)]$, where α denotes the phase

$$\Psi_{(C)}(\mathbf{r}_1, \mathbf{r}_2; t) = \frac{1}{2} [\chi_L(\mathbf{r}_1; t) + \exp(i\alpha) \chi_R(\mathbf{r}_1; t)] [\chi_L(\mathbf{r}_2; t) + \exp(i\alpha) \chi_R(\mathbf{r}_2; t)], \quad (7)$$

difference that evolved between the right and left wave functions during the adiabatic splitting of the wells. This two-particle, mutually coherent wave function takes the form where the label C stands for coherent. The mean field or classical description—see Equation (1)—of the double-well BEC assumes such mutual coherence. The single-particle density of the mutually coherent, freely expanding two-particle system reads

$$\rho_{1(C)}(\mathbf{r}; t) = \frac{1}{2} \left[\left| \chi_L(\mathbf{r}; t) \right|^2 + \left| \chi_R(\mathbf{r}; t) \right|^2 + \left(\chi_L(\mathbf{r}; t) \chi_R^*(\mathbf{r}; t) \exp(i\alpha) + \text{c.c.} \right) \right], \quad (8)$$

where c.c. is the complex conjugate of the previous term. Far from the potential minima of the initial wells, the amplitudes of the expanding wave functions vary slowly in space, so that we can approximate those amplitudes as $\chi_R(\mathbf{r}; t) \approx \chi \exp[i\theta_R(\mathbf{r}; t)]$ and $\chi_L(\mathbf{r}; t) \approx \chi \exp[i\theta_L(\mathbf{r}; t)]$, and the single-particle density in the far region becomes

$$\rho_{1(C)}(\mathbf{r}; t) = \chi^2 \{ 1 + \cos [\theta_R(\mathbf{r}; t) - \theta_L(\mathbf{r}; t) + \alpha] \} . \quad (9)$$

Thus, in addition to the densities of the expanding single-particle wave functions, $\rho_{1(C)}(\mathbf{r}; t)$ also contains an α -dependent term—namely, the interference fringes—that varies sinusoidally with the difference of the position-dependent phases of the overlapping χ_R and χ_L functions. The expression in Equation (8) is quite general; the single-particle density of an N -particle BEC distributed over two wells in a mutually coherent state takes on the form of Equation (9) in the far region.

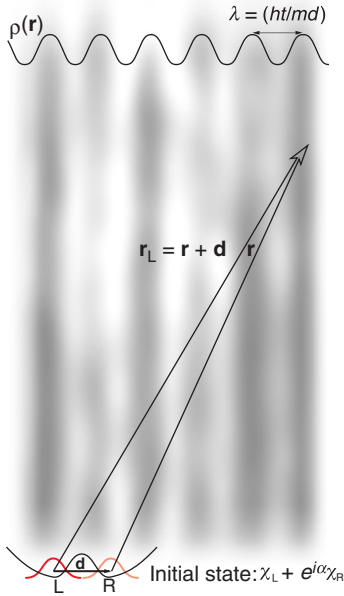


Figure 6. Geometry of Interference Fringes

The diagram shows the interference fringes in the image of two expanding BECs that were initially trapped in the right (R) and left wells (L) of a double-well potential. As defined in the text, the \mathbf{r} -vector denotes the position relative to the center of the right well, and the \mathbf{d} -vector denotes the relative position of the centers of both wells. The high-density regions of the interference fringes are planes oriented perpendicular to \mathbf{d} . At a time t after releasing the BECs, the interference fringe planes are separated by a distance $\lambda = \hbar t/(md)$. The actual positions of the fringes depend on the phase difference α of the initial BECs (if the BECs are phase coherent, $\chi = \chi_L + e^{i\alpha}\chi_R$).

Heuristic Derivation of the Interference Fringe Pattern. What is the geometry and spacing of the interference fringes that would be produced by this mutually coherent state? We offer a heuristic derivation of the phase of a freely expanding single-particle state. Classically, a particle that has traveled a distance r in a time t has a velocity $v = r/t$. In the spirit of the Madelung description, we associate the gradient of the phase θ with mv/\hbar , and we find $d\theta/dr = (mr/\hbar t)$, so that $\theta = (m/2\hbar)(r^2/t) + C$, where C denotes a constant, independent of \mathbf{r} . Now we suppose that the left and right BECs are sufficiently alike so that we can assume that their phases in the expansion evolve with the same constant C . In that case, the difference between the phases of the amplitudes χ_R and χ_L evaluated at a vector distance \mathbf{r} from the center of the right well and \mathbf{r}_L from the center of the left-well is

$$\theta_R - \theta_L = (m/2\hbar t)[r^2 - r_L^2] = - (m/2\hbar t)[2\mathbf{d} \cdot \mathbf{r} + d^2] , \quad (10)$$

where the vector distance \mathbf{d} separates the centers of the potential wells and $r^2 - r_L^2 = -2\mathbf{r} \cdot \mathbf{d} - d^2$ (see Figure 6). The high-density regions of the interference fringes are planes perpendicular to \mathbf{d} at a regular spacing of $\lambda = \hbar t/(md)$. The measured density pattern for the density in Equation (9) is

$$\rho_{1(C)}(\mathbf{r}; t) = \chi^2 \left\{ 1 + \cos \left[(m/\hbar t) (\mathbf{r} \cdot \mathbf{d}) + (m/2\hbar t) d^2 - \alpha \right] \right\} , \quad (11)$$

and the value of α can be inferred from the positions of the interference fringe planes. A more careful derivation of the phases $\theta_{R(L)}$ gives corrections, but the above expressions are essentially correct in the regions imaged in the interfering BEC experiment. The experimental images do indeed reveal planar interference fringes, separated by a distance $\lambda = \hbar t/(md)$.

Resolving the Origin of the Interference. The experiment clearly indicated coherence, and the image agrees with the single-particle density of the mutually coherent double-well system. However, the experimental system was prepared not in a mutually coherent state, but in a number state analogous to that described by Equation (5). In that state, given that the single-particle density $\rho_{1(N)}$ in Equation (6) does not exhibit interference, why does the recorded image show fringes like those from the coherent single-particle density in Equation (11). The resolution of this apparent puzzle lies in the fact that the image does not record the single-particle density. Instead, the experiment probes the multiparticle density. Specifically, we cannot interpret the image of the N -particle system as N independent measurements of the single-particle density. But we can assume that the measurement captures the N -body system in a “likely” configuration; that is, the observation of a particle at \mathbf{r}_1 , another at \mathbf{r}_2 , and so on, indicates that the state of the system corresponding to the N -particle density $\rho_N(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = |\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2$ has a relatively high probability.

We use the special case of two particles in a double-well potential to illustrate the difference in probing the N -particle rather than the single-particle density. We assume the two-particle double-well system is prepared in the number state of Equation (5). We detect the particles at a time t during a period that is short on the time scale on which the single-particle wave functions χ_L and χ_R expand. The probability that one particle is recorded at \mathbf{r}_1 and the other at \mathbf{r}_2 is proportional to the two-particle number-state density:

$$\rho_{2(N)}(\mathbf{r}_1, \mathbf{r}_2; t) = |\chi_L(\mathbf{r}_1; t)|^2 |\chi_R(\mathbf{r}_2; t)|^2 + |\chi_L(\mathbf{r}_2; t)|^2 |\chi_R(\mathbf{r}_1; t)|^2 + \left[\chi_L(\mathbf{r}_1; t) \chi_R^*(\mathbf{r}_1; t) \chi_R(\mathbf{r}_2; t) \chi_L^*(\mathbf{r}_2; t) + \text{c.c.} \right] . \quad (12)$$

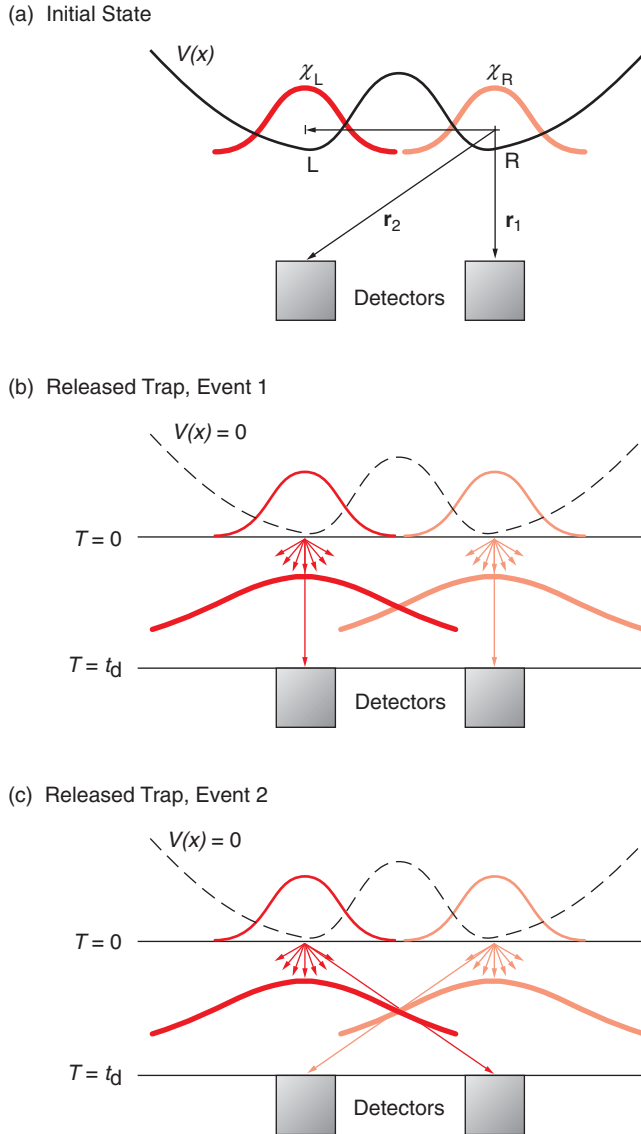


Figure 7. Origin of the Two-Particle Interference in Equation (13)

This schematic illustrates the origin of the interference pattern in the two-particle density for an expanding two-particle system that originates in a number state of a double-well potential. (a) The origin of the coordinate system is the center of the right potential well. (b) In event 1, the particle detected at r_1 originates from the right well; the particle detected at r_2 , from the left well. (c) In event 2, the particle detected at r_1 originates from the left well, whereas the particle detected at r_2 originates from the right well.

Because the two-particle wave function consists of a superposition of terms that correspond to the classical trajectories shown in (b) and (c), these events can interfere.

Assuming that \mathbf{r}_1 and \mathbf{r}_2 are located in the region where $|\chi_L(\mathbf{r}_1; t)| \sim |\chi_L(\mathbf{r}_2; t)| = \chi$, the two-particle density defined in Equation (12) takes on the form

$$\rho_{2(N)}(\mathbf{r}_1, \mathbf{r}_2; t) \approx \chi^4 \left\{ 1 + \cos \left[\theta_R(\mathbf{r}_1; t) - \theta_L(\mathbf{r}_1; t) - (\theta_R(\mathbf{r}_2; t) - \theta_L(\mathbf{r}_2; t)) \right] \right\}, \quad (13)$$

which contains the typical oscillatory contribution seen in Equation (9) describing an interference pattern. Thus, although the system is in a number state and the single-particle density does not exhibit interference, the two-particle density $\rho_{2(N)}$ does show interference.

The sinusoidal contributions in Equation (13) arise from the interference of the two distinct two-particle events illustrated in Figure 7. In one event, the particle detected at \mathbf{r}_1 was initially in the right well, whereas the particle detected at \mathbf{r}_2 originated from the left well. In the second event, the situation is reversed: The particle detected at \mathbf{r}_1 originated from the left well, whereas the particle detected at \mathbf{r}_2 originated from the right well.

Using Equation (10) for the phase difference between the two single-particle wave functions at a position \mathbf{r} , $\theta_R(\mathbf{r}) - \theta_L(\mathbf{r}) \approx -(m/2\hbar t)[2\mathbf{d} \cdot \mathbf{r} + d^2]$, we find that the two-particle distribution depends only on the relative position $\mathbf{r}_1 - \mathbf{r}_2$,

$$\rho_{N,2}(\mathbf{r}_1, \mathbf{r}_2; t) \approx \chi^4 \{1 + \cos[(m/\hbar t)\mathbf{d} \cdot (\mathbf{r}_1 - \mathbf{r}_2)]\} . \quad (14)$$

Whereas the likelihood of detecting the first particle at position \mathbf{r}_1 is independent of \mathbf{r}_1 in the far region $\rho_{1(N)} \approx \chi^2$, the likelihood of detecting a second particle at \mathbf{r}_2 is greater near the planar regions $\mathbf{d} \cdot (\mathbf{r}_1 - \mathbf{r}_2) = n(\hbar t/m)$, where n denotes an integer. Note that the planar regions of maximal $\rho_{2(N)}$ -values resemble the interference fringes of $\rho_{1(C)}$ in Equation (11), namely, the single-particle density of the expanding, mutually coherent two-particle double-well system. In fact, the fringe patterns for the two-particle density will be identical to those of an equivalent mutually coherent system, provided the relative phase α is chosen so that the fringes of that equivalent system overlap the position where the first particle was detected. Because the position of the first particle is undetermined until measured, we can say that it is the act of determining the first particle's position that fixes the value of the relative phase of an equivalent mutually coherent system. The two-particle number-state probability distribution then resembles the product of one-particle probability distributions of the equivalent mutually coherent system. That equivalence is a general feature: The more particles detected in the image of an expanding number-state double-well BEC, the more the outcome of such measurement resembles that performed on a mutually coherent double-well BEC. The relative phase of the equivalent mutually coherent BEC system can be extracted from the image but cannot be determined beforehand.

The equivalence to a mutually coherent state with a value of the phase difference that is established by the act of measurement is familiar from the observation of interference of independent lasers (Pfleeger and Mandel 1967) and of the dc Josephson effect (Anderson 1986).

Relative Phase Dynamics for Two N -Particle BECs. Our derivation of the number-state two-particle density and its equivalence to a mutually coherent state density of undetermined relative phase is not easily generalized to a number-state double-well system with larger particle numbers. Instead, we can apply the elegant description developed for the relative phase dynamics of Josephson junctions. In this description, the dynamics between the two weakly linked superfluids is cast in terms of only two variables: α , the relative phase, and m , half the difference of the number of particles contained in each well. In fact, m and α are quantum numbers, and the number states are the eigenstates of m . We denote by $|m\rangle$ the number state of a double-well system with N -particles per well, of which $N - m$ occupy the left well and $N + m$, the right well.

An alternative set of basis functions is provided by states of good relative phase $|\alpha\rangle = N^{-1/2} \sum_m \exp(i\alpha m) |m\rangle$. The transformation from the $|m\rangle$ -basis to an $|\alpha\rangle$ -state representation is therefore a Fourier transform, somewhat analogous to the transformation between the traditional momentum and coordinate representations. Just as coordinates and momenta are conjugate to each other, m and α are conjugate variables. The many-body state can be expanded in either the $|\alpha\rangle$ -states or the $|m\rangle$ -states, $|\Psi\rangle = \int d\alpha \Psi(\alpha) |\alpha\rangle = \sum_m \Psi_m |m\rangle$, where $\Psi(\alpha)$ and Ψ_m are equivalent to the coordinate (x) and momentum (p) representations of a single-particle state. Generally, the Ψ wave function implies a spread both in the m and α variables: $\Delta m = (\langle (m - \langle m \rangle)^2 \rangle)^{1/2}$, $\Delta \alpha = (\langle (\alpha - \langle \alpha \rangle)^2 \rangle)^{1/2}$, where $\langle \rangle$ denotes the expectation value. As conjugate variables, Δm and $\Delta \alpha$ satisfy the

Heisenberg uncertainty relation $\Delta m \times \Delta \alpha \geq 1$, whereas Δx and Δp satisfy the relation $\Delta x \times \Delta p \geq \hbar$ in single-particle quantum mechanics.

To continue our comparison of BEC interference experiments with single-particle quantum mechanics, we note that the establishment of a relative phase between interfering BECs is the analogue of a position measurement on a particle in a plane-wave state. When the plane wave has a well-defined momentum, then $\Delta p = 0$ and $\Delta x \rightarrow \infty$. The latter expression means that the coordinate has maximum uncertainty, and therefore, a measurement of x could yield any value. Likewise, in the initial state of the interfering BEC experiment, $\Delta m = 0$, and the determination of α achieved by the imaging of the expanding BECs could yield any value. When the measurement is performed, however, the wave function collapses to an eigenstate of α .

Squeezing the Numbers in BECs—Macroscopic Quantum Fluctuations

As mentioned previously, the number-phase description in terms of the α or m quantum eigenvalues is familiar from the treatment of Josephson junctions. The application of the number-phase description to the problem of double-well BECs then reveals an intimate connection between the physics of BEC interference and Josephson physics. However, the BEC interference experiment conducted at MIT lacks the weak link through which the superfluids can exchange their boson particles. Consequently, it is not exactly a BEC-Josephson experiment. In a subsequent effort, the Kasevitch group at Yale used a related setup and succeeded in inducing and controlling such reversible superflow between multiple BECs. The Yale experimentalists achieved this goal by trapping the BECs in the potential minima of an optical lattice—a trapping potential that oscillates sinusoidally in space as $|E_0|^2 \sin^2(kx)$ —and by lowering and raising the potential barriers separating the BECs through variations of $|E_0|^2$. Most important, the Yale group probed Josephson physics by observing variations in the interference pattern of the expanding BECs after switching off the optical-lattice potential. The sharpness of the interference fringes revealed the uncertainty in relative phase, $\Delta \alpha$, of the expanding BECs. In particular, when the barrier height had been sufficiently increased before the BECs were released, the fringes observed in the image of the expanding BECs became fuzzy, an indication that the uncertainty in the phase values of the initial BECs had increased markedly. This increase is expected as the number uncertainty decreases. As we argue below, this is a genuine quantum fluctuation effect observed in a macroscopic system. To set the stage, we start by elucidating the role of the quantum fluctuations in multiple-well BEC physics.

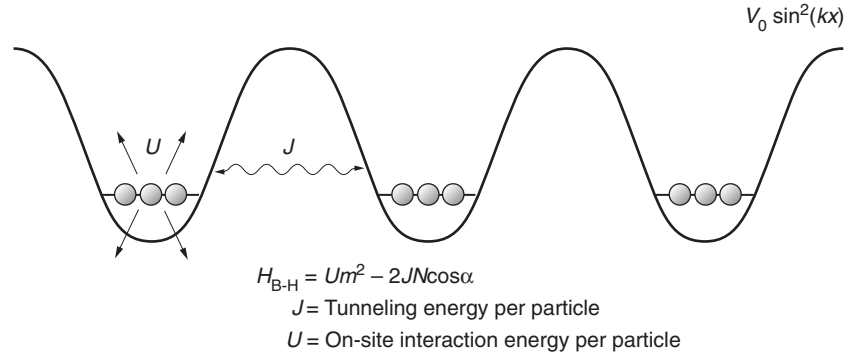
Quantum and Classical Physics of Double-Well BECs. As in Equation (1), the classical or mean-field description of the N -particle double-well system, the many-body wave function is a product state: $\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; t) \approx \chi(\mathbf{r}_1; t) \dots \chi(\mathbf{r}_N; t)$, where each single-particle wave function is a linear superposition of left-well (χ_L) and right-well (χ_R) wave functions,

$$\chi(\mathbf{r}; t) = 1/(2N)^{1/2} [(N - m(t))^{1/2} \chi_L + e^{i\alpha(N + m(t))^{1/2}} \chi_R] , \quad (15)$$

and α and m are well-defined parameters. We use the same notation as in the number-phase description because the physical interpretation of α and m is the same as that of the quantum eigenvalues introduced above. In fact, $\alpha(t)$ and $m(t)$ in Equation (15) are the expectation values of the quantum treatment of the number-phase dynamics. The classical treatment can then describe superfluid effects, the essence of which relies on

Figure 8. The Bose-Hubbard Model

The diagram shows an optical-lattice potential occupied by atoms of integer spin. The interactions between the atoms include a hopping or tunneling interaction and a repulsive interaction between atoms at the same site.



the existence of a well-defined phase—see Equations (1) to (4) and the section “The Coherent Wave Nature of Superfluidity”—but it cannot account for behavior such as the collapse to a random value of the relative phase in the imaging of interfering BECs. More generally, contrary to predictions of classical mechanics, the quantum treatment predicts different outcomes of identical measurements on identically prepared systems. Measures of such quantum randomness are the standard deviations, such as the deviations $\Delta\alpha$ and Δm introduced earlier, that quantify the range of quantum fluctuations. For sufficiently large numbers of atoms, Δm can take on values that are large enough for the fluctuation range to be called “macroscopic.”

Weakly Linked BECs. When the barrier separating the two potential wells in the double-well BEC is lowered to an appropriate value, atoms can penetrate the barrier, which thereby provides the weak link that allows the left and right BECs to exchange particles. As in the description of BEC interference, we define a phase for each BEC and describe the possible particle exchange in terms of the canonically conjugate variables that represent the difference of the condensate phases, α , and half the difference of the particle population, m , occupying the individual BECs. The inter-BEC particle exchange gives rise to an effective tunneling energy of the usual Josephson form,

$$E_{\text{tun}} = -E_J \cos(\alpha) . \tag{16}$$

We expect the value of E_J to be roughly proportional to the number of particles (N) per well, to depend weakly on the number difference m , and to be extremely sensitive to the height of the potential barrier separating the BECs. As the barrier height increases, the tunneling of particles is restricted, a limitation corresponding to a decrease in the value of the E_J -parameter in Equation (16). In what follows, we write $E_J = 2NJ$, where J denotes the tunneling energy per particle. The tunneling energy, minimized by putting $\alpha = 0$, favors a well-defined value of the phase difference in the ground state and, hence, favors the establishment of a definite phase difference (the superfluid limit). In contrast, the usual interparticle interactions, if repulsive, favor a well-defined value of m . To see that, we note that the interparticle interaction energy scales as the number of interactions. The N_L -particles (in the left BEC) experience $N_L(N_L - 1)/2 \approx N_L^2/2$ interactions. Similarly, the N_R -particles (contained in the right well) undergo $N_R^2/2$ interactions. Assuming that the interaction energy per particle, U , is approximately the same in each well and using $N_L = N - m$ and $N_R = N + m$, we write the total interaction energy as

$$E_{\text{int}} = (U/2)[N_{\text{R}}^2 + N_{\text{L}}^2] \approx U[N^2 + m^2] . \quad (17)$$

In contrast to the tunneling energy, E_{int} takes on its minimum value at $m = 0$, corresponding to the BEC number state with $N_{\text{R}} = N_{\text{L}} = N$. The contribution to the energy that stems from the phase-number dynamics (the sum of interaction and tunneling energies after the constant UN^2 -term has been discarded) is then equal to

$$H = Um^2 - 2JN \cos(\alpha) . \quad (18)$$

Classically, the position of lowest energy is $m = 0$, $\alpha = 0$. Quantum mechanically, it follows from Heisenberg's uncertainty principle that m and α , being conjugate variables, cannot be determined simultaneously to absolute certainty. We now use the double-well Bose-Hubbard Hamiltonian in Equation (18) as a starting point to indicate how weakly linked BECs can be regarded as a laboratory for exploring both the classical dynamics and the quantum nature of Josephson junctions. A schematic representation of the individual terms that contribute to this Hamiltonian is shown in Figure 8.

Probing Josephson Physics in Weakly Linked BECs. The Bose-Hubbard Hamiltonian in Equation (18) is the generic form of the Hamiltonian that governs the physics of Josephson junctions. We can expect, therefore, that the atom trap becomes a new laboratory for studying Josephson effects. Although this physics has been studied intensely in condensed-matter environments, the new parameter range and technology of the BEC traps give a new twist to the study of Josephson-junction physics and other known phenomena, as well as the opportunity to study quantum fluctuations and, perhaps, to discover novel applications.

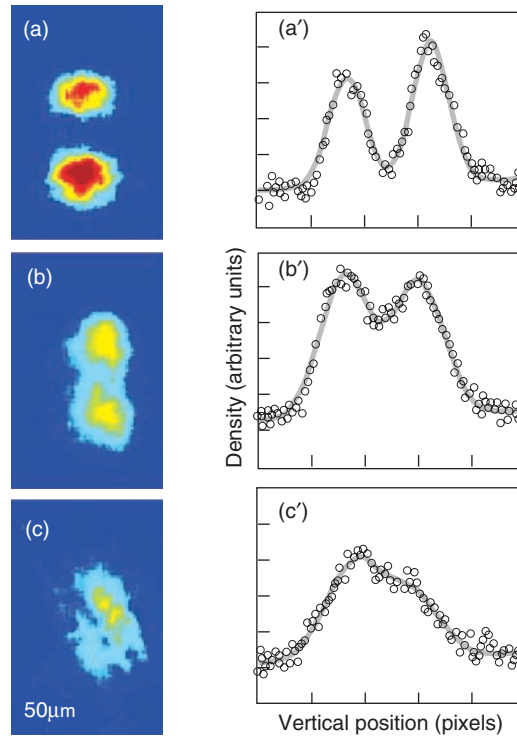
A sudden change in the depth of one of the wells or in its particle number can “nudge” the many-body system out of equilibrium, initiating a collective excitation in which the expectation value of the well populations oscillates. This phenomenon is called Josephson oscillations. On the topic of probing quantum behavior, it is interesting that the parameters in Equation (18) can be controlled experimentally: Variations in the trapping potential can alter the values of U and J . Clearly, the atom-trap technology gives unusual control over the Josephson junction, providing new knobs that can both initiate Josephson oscillations and vary the quantum fluctuations. The crucial question of whether oscillations and fluctuations can be measured in cold-atom BECs was answered, in part, by the Yale experiment.

What are the obstacles that the BEC technology faced in probing Josephson physics? In superconductors, Josephson effects are routinely studied by measurements of the weak supercurrent. Such measurement of a charged particle can be achieved relatively simply and accurately. In systems of neutral particles, on the other hand, the observation of a weak current represents a much greater challenge, and in helium fluids, a Josephson current was only recently observed (Packard 1998). By the same token, in the neutral-atom traps, current atom-counting techniques are not sufficiently accurate to allow observing small-amplitude population oscillations. Numbers of atoms in a typical BEC are measured with a relative accuracy of only about 10 percent. This low accuracy renders the technique unsuitable for observing Josephson oscillations of atomic-trap populations in the linear regime (number oscillations with a magnitude of 1 percent or less of the total number of trapped atoms). Instead of measuring a population imbalance,

Figure 9. Formation of Number-Squeezed States in an Atom-Trap BEC

The sequence of absorption images, (a)–(c), and the associated density cross sections, (a')–(c'), show atoms released from optical lattices of increasing depth: $U_0 = 7.2E_{\text{recoil}}$, $U_0 = 18E_{\text{recoil}}$, and $U_0 = 44E_{\text{recoil}}$, respectively. In (a), the two-peaked structure is due to interference between atoms released from different lattice sites. As the well depth increases and the tunneling rate decreases, the interference pattern becomes progressively blurred, reflecting greater phase uncertainty and the formation of number-squeezed states.

(Reprinted with permission from Orzel et al. *Science* 291, page 2389. Copyright 2001 American Association for the Advancement of Science.)



we might try to observe the relative phase of BECs, which gives a complementary view of the physics; for instance, the expectation value of the relative phase oscillates at the same frequency as the population imbalance or current in the Josephson oscillation. The BEC interference experiment conducted at MIT illustrated that the relative phase can be measured from recorded images of expanding BECs. This measurement, however, is destructive and yields a value for the phase at a single time. Whether this technique could be used to probe the time evolution of the phase is not evident. In addition, the imaging of BEC interference in the double-well system gives only a single value of the phase, whereas a measurement of the range of quantum fluctuations requires a record of the phase distribution.

The Yale experiment resolved the problem of probing the phase distribution by imaging the interference of many simultaneously expanding BECs, which had been weakly linked before the trapping potential was released. The resulting image is sensitive to the distribution of the complex phase values of the BECs. If the phases of the BECs are strongly correlated—they all have approximately the same value, for instance—then the interference of each pair of BECs can add up in phase and give an overall pattern with bright and sharp fringes. In contrast, if the phases of the weakly linked BECs are randomly distributed, then their values, determined by the act of imaging, differ widely. As a consequence, the fringes corresponding to the interference of different pairs of BECs do not overlap, so that interference washes out. The Yale experiment imaged the density of 12 expanding BECs that had been initially trapped in the adjacent potential wells of a linear optical lattice and weakly linked before the optical-lattice potential was released (see Figure 9). In such an optical lattice, the centers of mass of adjacent BECs are all separated by the same distance (half the wavelength of the light that creates the standing wave pattern of the lattice potential). By measuring the amplitude and fringe sharpness (defined as the ratio of spatial width to the distance separating the fringes) observed in imaging the expanding BECs, the Yale group quantified the uncertainty of the relative phase values.

As they had ramped up the height of the potential barriers before releasing the BECs, the Yale group observed a marked decrease in the sharpness of the fringes in the expanding-BEC images. The measured sharpness was in quantitative agreement with numerical simulations that were based on the ground-state phase uncertainty. The assumption that the many-body system has reached its ground state before the trapping potentials are switched off is reasonable because the change in potential barrier was effected adiabatically in the experiments. In a ground state, the uncertainties of conjugate variables generally reach the Heisenberg limit, which in this case would mean that $\Delta m \times \Delta \alpha \approx 1$. Thus, from their measurements and the agreement with the predicted values of phase uncertainty, the Yale group inferred that their observed

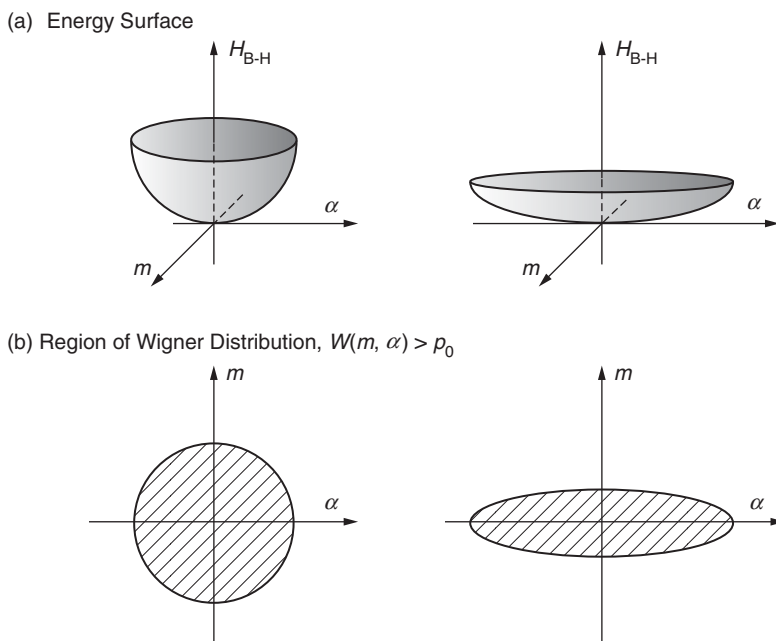


Figure 10. Number Squeezing in Phase Space

This graph illustrates the physics of number squeezing by showing the effect of an increase in the potential barrier on the number phase (m, α) Wigner distribution of the double-well BEC discussed in the text. The graphs show the area in which the Wigner distribution of the many-body ground state exceeds a minimal value. An increase in the potential barrier lowers the tunneling rate J , which reduces the {tightness of confinement in the α -direction of the (m, α) phase space. The word “confinement” refers to the potential energy-like term in the energy expression of Equation (18) that depends on α . As a result of lowering J , the ground-state Wigner distribution stretches out in the α -direction. In accordance with the Heisenberg uncertainty principle ($\Delta m \Delta \alpha \approx 1$), the area of high Wigner distribution value remains constant in the process of stretching and the number uncertainty Δm decreases accordingly.

increase in phase uncertainty implied a similar decrease in number uncertainty Δm . By analogy with a similar reduction of uncertainty in optical field intensities, the process of reducing $\Delta m \ll N^{1/2}$ is called “squeezing.” In Figure 10, we further illustrate the aptness of this term by sketching the effect of varying the parameters of the Hamiltonian in Equation (18) on the Wigner distribution function.

The experimental increase of the potential barrier height lowers the value of J , which greatly reduces the tightness of the confinement in the α -direction of the (α, m) -phase space. In response, the Wigner distribution stretches out farther in the α -direction. Since the area of high probability shown in Figure 10 remains of order 1, the uncertainty in the m -direction is tightly squeezed. Thus, as the hopping motion of particles between adjacent wells is “frozen out,” each well contains a better-defined number of particles. To further support their claim of having observed quantum fluctuations, the Yale group also demonstrated that the trend of decreased fringe sharpness may be turned around by reversal of the variation in potential barrier height.

Quantitative Treatment of Number Squeezing. We now revisit the description of the double-well BEC to provide a quantitative understanding of the number uncertainty squeezing illustrated in Figure 10. We introduce a dimensionless parameter, or coupling constant Γ , that characterizes the competing interactions in the system: $\Gamma = UN/2J$ is the ratio of the interparticle interaction energy per well ($UN^2/2$) to the tunneling energy per well NJ (the latter plays a role somewhat analogous to that of kinetic energy in other systems). We minimize the Hamiltonian described by Equation (18) in the α -representation. To convert Equation (18) from the number representation to the α -representation, we replace the m -operator by $-(1/i)\partial/\partial\alpha$. Then, we calculate the expectation value of the Hamiltonian by using the Gaussian state for the wave function ψ , $\psi(\alpha) \propto \exp(-\alpha^2/(4x))$. The expectation values are simplified when expressed in terms of the width parameter x , which is related to the uncertainty in phase difference as $\Delta\alpha = (2x)^{1/2}$: $\langle m^2 \rangle = -\langle \partial^2/\partial\alpha^2 \rangle = 1/(2x)$ and $\langle \cos(\alpha) \rangle = \exp(-x)$.

The expectation value of the Hamiltonian is then equal to

$$\langle H \rangle = \frac{U}{2} \left[\frac{1}{x} - \frac{2N^2}{\Gamma} \exp(-x) \right] \quad (19)$$

and we obtain the value of the width parameter x by minimizing Equation (19):

$$x = \sqrt{\frac{\Gamma}{2N^2}} \exp(x/2) . \quad (20)$$

In the weakly coupled regime $\Gamma \ll 2N^2$, the minimum expectation value occurs at a value $x \ll 1$, in which case $\exp(x/2) \approx 1$ and Equation (20) yields the width parameter $x \approx (\Gamma/2)^{1/2}/N$. In other words, the weakly coupled case yields a very small phase uncertainty,

$$\Delta\alpha = (2x)^{1/2} \approx (2\Gamma)^{1/4}/N^{1/2} \ll 1 , \quad (21)$$

and therefore corresponds to the superfluid limit. Most superconducting Josephson junctions find themselves in this limit. Because the number uncertainty is small, $\Delta\alpha \sim N^{-1/2}$, the classical (or mean-field) approximation successfully describes these Josephson experiments. The uncertainty in particle number $\Delta m \approx N^{1/2}/2\Gamma^{1/4}$ appears Poissonian ($\Delta N \approx N^{1/2}$) if we write it in terms of the coupling constant. The small phase uncertainty in this regime is not easily measured with appreciable accuracy.

In contrast, as the value of J is lowered by an increasing barrier height, the coupling constant $\Gamma = [UN/2J]$ increases accordingly, and the phase uncertainty can increase to give a measurable decrease in fringe sharpness. In the Yale experiment, the increase in the potential barrier was sufficient to allow the system to approach the strong coupling regime $\Gamma \sim N^2$ or $U/J \sim N$. In that regime, the value of x at the minimum energy can become of order 1, in which case we cannot replace $\exp(x/2)$ by 1. Instead, we must solve Equation (20). By the time the potential barrier has been increased to the point that, say, $U/J = (4/e)N$, the variation becomes $\Delta m = (1/2)^{1/2}$, and the uncertainty in atomic population of each well has dwindled to less than one particle. At that point, $\Delta m \ll N^{1/2}$, and we say that the number distribution has become sub-Poissonian. The phase-difference uncertainty, $\Delta\alpha$, also becomes of order unity. Well before that point, say, when U/J is increased to only 10 percent of N , or $U/J = 0.1N$, the uncertainty in phase difference in the double-well BEC has grown to half a radian. In the multiple-well BEC system, the uncertainty in phase between nonadjacent wells under that same condition, $U/J = 0.1N$, is greater, and the loss of fringe sharpness in the interference of 12 BECs is quite noticeable.

From Superfluid to Mott Insulator

By illustrating number squeezing, the Yale group demonstrated that BEC technology can engineer and observe quantum fluctuations of an almost macroscopic system. On the other hand, technical constraints in the Yale experiment limited the height to which the potential barrier could be raised and, hence, the range to which the number uncertainty could be squeezed. These limitations prevented the Yale group from venturing further into the strong-coupling regime. By pushing this frontier, Hänsch's group in Munich were able to observe a very interesting phase transition (Figures 11 and 12). As they squeezed the number uncertainty below a value of order 1—it would be $(1/2)^{1/2}$

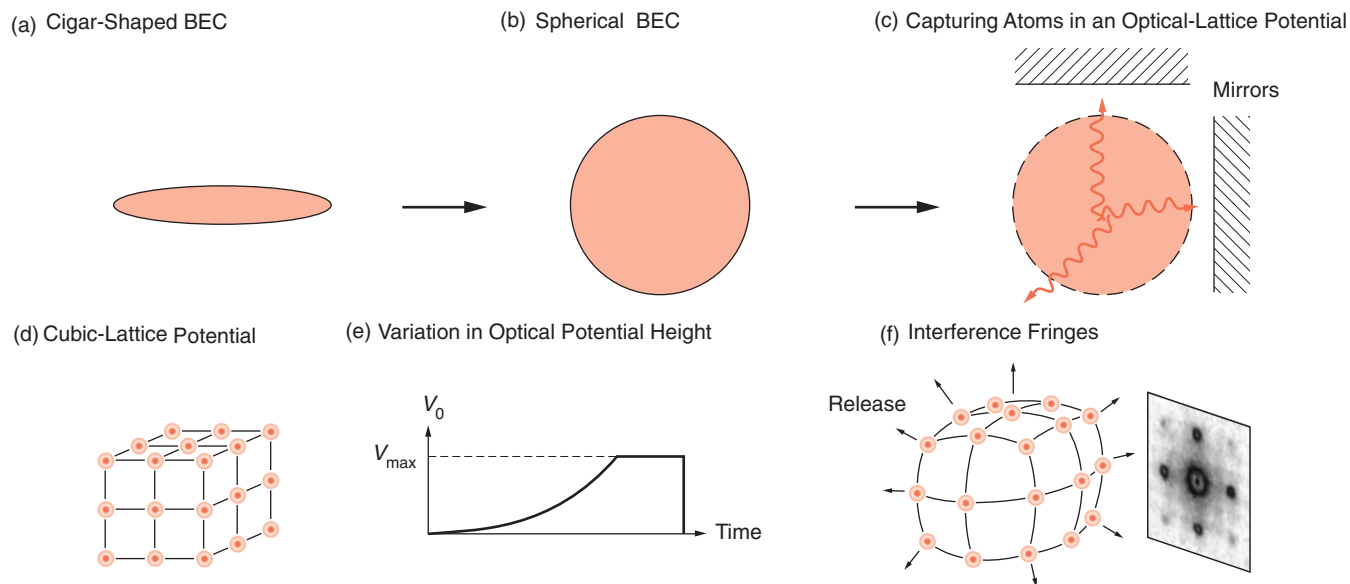


Figure 11. Demonstration of a Transition from a Superfluid to a Mott Insulator

In the BEC experiment that demonstrated the quantum phase transition from a superfluid to a Mott insulator (Greiner 2002), the experimentalists started with a cigar-shaped BEC (a) that was relaxed to a spherical BEC (b), distributing the atoms more evenly over a larger region of space. By shining in three laser beams, detuned from each other and reflected by mirrors, the researchers created a standing-wave pattern that captures the atoms in an optical-lattice potential (c): $V(x, y, z) = V_0[\sin^2(kx) + \sin^2(ky) + \sin^2(kz)]$, where k denotes the wave vector of the laser light. Gradual increases in laser

intensity and, hence, in the potential V_0 trap one to three atoms per potential minimum, or well. These minima form a cubic lattice (d). In (e) we show a typical variation of the optical potential height V_0 : The potential height is ramped up “slowly” for 80 ms and kept constant for another 20 ms; then the trapping potential is suddenly switched off, at which point the atoms in the BEC begin to expand. In (f), the atomic wave functions from different wells begin to overlap, and the atomic density imaged in a plane shows interference fringes.

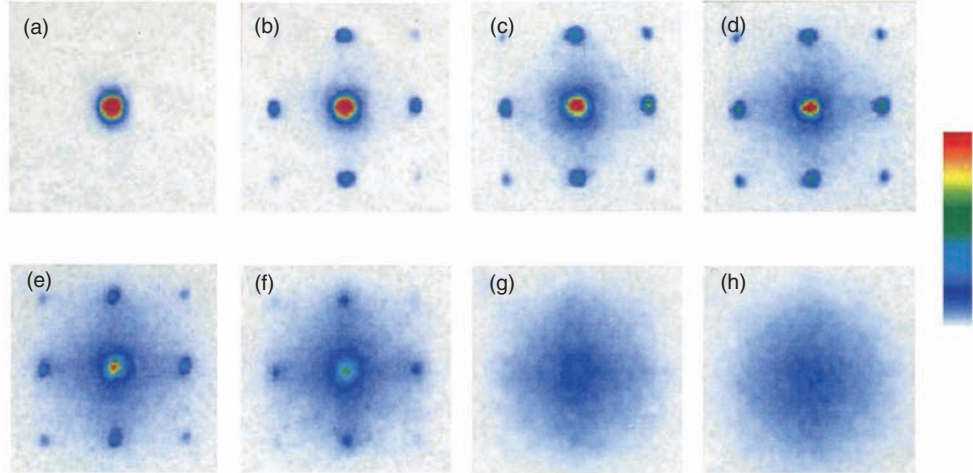
in the approximations introduced previously—the ground state abruptly changes to a Fock or number state with $\Delta m = 0$. This phenomenon is a true phase transition: Many-body properties change suddenly as $U/J \sim N$. In addition to the change in number statistics, the system’s conductivity alters discontinuously as the system takes on a number state. In the number state, a finite amount of energy is required to transfer atoms between wells; therefore, the transition to the number state abruptly alters the nature of the many-body system from a conductor with superfluid properties to a Mott insulator. This many-body phenomenon is an example of a transition driven by the competition between different interactions, rather than by the competition between order and disorder, which is responsible for usual phase transitions. If they involve quantum fluctuations, the former transitions (which occur at zero temperature) are called quantum phase transitions.

If we can trust the tunneling energy in Equation (16) and the interaction energy in Equation (17) to accurately describe the many-body physics, then the BEC in an optical lattice is an example of a Bose-Hubbard system. The theory of the phase transition from superfluid to Mott insulator in such systems has been explored in great detail. Experimentally, this transition was first observed in an array of superconducting Josephson junctions. In BEC physics, the experimental study of the transition by the Munich group demonstrated, once again, that the BEC technology gives an unusual degree of control.

Figure 12. Absorption Images Showing a Transition to a Mott Insulator in a BEC

The BEC absorption images (a)–(h) were recorded in a particular plane 15 ms after the trapping potential was switched off. The images reflect different maximum values V_{\max} of V_0 . In units of the recoil energy, $E_{\text{recoil}} = \hbar^2 k^2 / 2m$ (capital R was used for “right”), V_{\max} took on the values (a) 0, (b) $3E_{\text{recoil}}$, (c) $7E_{\text{recoil}}$, (d) $10E_{\text{recoil}}$, (e) $13E_{\text{recoil}}$, (f) $14E_{\text{recoil}}$, (g) $16E_{\text{recoil}}$, and (h) $20E_{\text{recoil}}$. Notice that the interference pattern completely disappears between $V_0 = 14E_{\text{recoil}}$ and $V_0 = 16E_{\text{recoil}}$, in agreement with the prediction that all phase information would be lost as the potential barriers increase and the atoms become localized in their respective potential wells.

(This figure was reproduced courtesy of *Nature*.)



Before describing the experiment, we demonstrate the transition in the double-well BEC system. From Equation (19), we see that, in the limit of large phase uncertainty ($x \rightarrow \infty$), the expectation value of the number-phase energy—Equation (18)—vanishes. Consequently, when the local minimum of $\langle H \rangle$ takes on a positive value, the true minimum of the system is found at $x \rightarrow \infty$, as we illustrate in Figure 13. As the value of U/JN increases, the value of the local minimum increases until, at $U/J = (4/e)N$, corresponding to $\Delta\alpha = 2^{1/2}$ and $\Delta m = 2^{-1/2}$, the value of the minimum turns positive and the real minimum is at $x \rightarrow \infty$, corresponding to $\Delta\alpha \rightarrow \infty$ and $\Delta m = 0$.

A significant difference between the Yale and Munich experiments lies in the number of potential wells created in the optical lattices. The trapping potential in Hänsch’s group was a three-dimensional lattice of 65 sites in each dimension. The large number of lattice sites in the Munich experiment, 65^3 in total, is significant because it allows experimentalists to trap one to three particles per site while still having a sufficiently large total number of atoms to image the interference of the expanding BECs. By lowering the value of N (N was about 10,000 in the Yale experiment), the Munich group could reach the critical ratio of $U/J \sim N$ with a much smaller increase in barrier height. Actually, the simple (α, m) treatment of the number-phase dynamics in the double-well BEC becomes invalid for small values of N and a different description, such as the one presented by Subir Sachdev (1999), is necessary. Nevertheless, the (α, m) description still captures the main features and predicts the correct order of magnitude of the transition point. Hänsch’s group also probed the excitations of this system and found evidence for the insulator property of a finite energy (or “gap”) necessary to allow transferring atoms between wells. Again, these experiments illustrate the unprecedented tools offered by the cold-atom technology.

Los Alamos Achievements and Future Work

With regard to fundamental physics, we have shown that BEC experiments can probe beyond the confines of traditional condensed-matter Josephson-junction studies by exploring and engineering quantum fluctuations. We have also emphasized that atom-laser systems with superfluid properties (long-range phase coherence in an equilibrium as opposed to a nonequilibrium state) may offer unique opportunities for application. For instance, the BECs may find novel uses in atom interferometry and sensing applications.

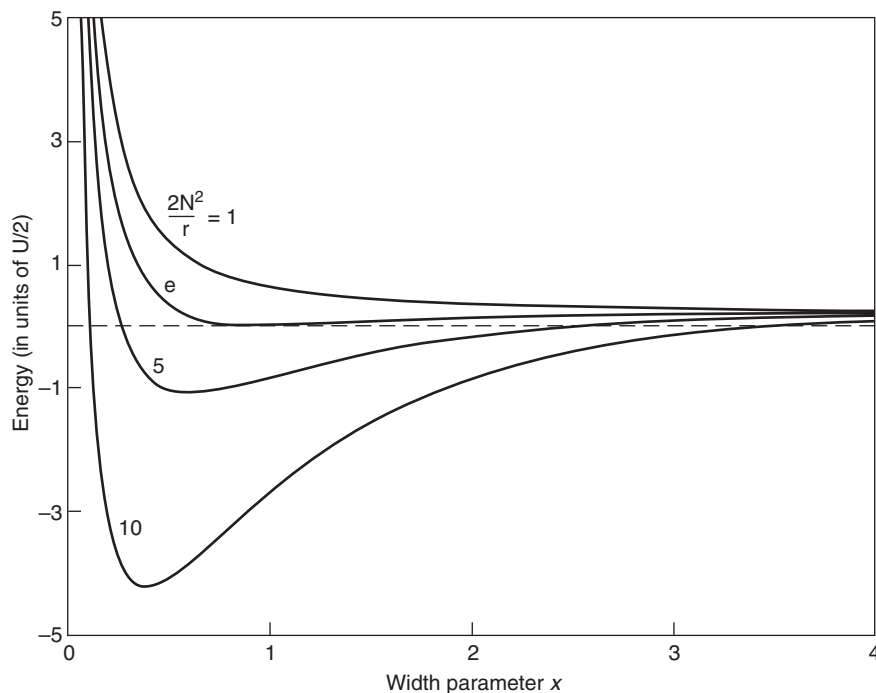


Figure 13. Number-Phase Energy for Different Interaction Parameters
 The expectation values of the number-phase energy of Equation (18) are calculated with a Gaussian trial wave function $\psi(\alpha) \propto \exp(-\alpha^2/[4x])$ and are plotted as a function of the width parameter x , which is related to the phase uncertainty $\Delta\alpha$ as $x = (\Delta\alpha)^2/2$. The different curves show $H(x)$ for different values of the interaction parameter $(2N^2/I)$. From bottom to top, those values are 10, 5, e , and 1. For $(2N^2/I) < e$, the local minimum is also the global minimum, whereas for $(2N^2/I) > e$, the global minimum occurs in the limit $x \rightarrow \infty$, corresponding to a complete uncertainty of the phase.

Hopefully, this historical perspective has also conveyed a sense of the flexibility of the cold-atom-trap technology. That flexibility has led to a host of other avenues being pursued or contemplated: for instance, schemes to alter and control the nature and strength of the interparticle interactions, already successful searches for superfluid properties in BECs, demonstrations of nonlinear physics effects in superfluids (vortices, solitons, and “quantum shocks”), the study of mutually coherent BECs, the demonstration of atom-molecule BECs, and the prospect of using BECs for the study of quantum measurement theory.

Los Alamos National Laboratory has been active in exploring several of the above aspects. The following are some of the Los Alamos contributions and ongoing projects that we are aware of. On the experimental side, David Vieira and Xinxin Zhao are working toward the use of an atomic BEC to cool down fermion atoms (see the article “Experiments with Cold Trapped Atoms” on page 168). On the theoretical side, Peter Milonni was the first to point out that external electric fields can be used to control the interparticle interactions in the atom-trap systems (Milonni 1996). Diego Dalvit, Jacek Dziarmaga, and Wojciech Zurek resolved the puzzle of the lifetime of the proposed Schrödinger cat states in BEC-like systems, and they have proposed schemes to reduce the effect of decoherence and increase the cat’s longevity (see the article “Schrödinger Cats in Atom-Trap BECs” on page 166). In collaboration with experimentalist Roberto Onofrio (visiting from the University of Padua, Italy), they continue to explore the possible use of BECs in studies of measurement theory. Lee Collins has explored the vortex and soliton dynamics in BECs, working closely with the experimental group of Bill Philips at the National Institute of Standards and Technology (Denschlag et al. 2000). Gennady Berman and Augusto Smerzi are exploring the possibility of using BECs to study the boundary between quantum and classical behavior (Berman et al. 2002), as well as using BECs in optical lattices for interferometry purposes (Dziarmaga et al. 2002).

Since 1996, I have also been active in BEC research. The prediction for the phase separation of BECs under specific conditions (Timmermans 1998) has been confirmed by experiments in Ketterle’s group at MIT. This same group also confirmed our

predictions for the reduction of scattering slow distinguishable particles by the BEC (Timmermans and Côté 1998) and for the excitation rate of phonon modes in two-photon scattering experiments. Recently, the group of Wieman at JILA found evidence for a prediction by Timmermans et al. (1999) of the formation of an atom-molecule BEC in the Feshbach resonance scheme that was initially proposed to alter the effective interparticle interactions. In a recent collaboration with Milonni of Los Alamos and Arthur Kerman of MIT, I pointed out the possibility of creating a fermion-boson superfluid (Timmermans et al. 2001) by bringing an ultracold fermion gas mixture near a Feshbach resonance. Finally, I discovered the heating mechanism that explains the temperature limit encountered by efforts in fermion atom cooling and provides the main obstacle for the current experiments to reach fermion superfluidity in atom traps (Timmermans 2001a).

The variety of approaches and cold-atom research topics at Los Alamos is yet another measure of the richness of this field. By now, numerous experiments have established the cold-atom trap as a new kind of laboratory in which to study interesting fundamental issues in low-temperature, many-body, and nonlinear physics. The unusual control and the variety of experimental knobs also hint at the possibility of practical applications. Hopefully, Los Alamos can continue to play a significant role in the ongoing cold-atom physics adventure. ■

Further Reading

- Anderson, P. W. 1986. Measurement in Quantum Theory and the Problem of Complex Systems. In *The Lesson of Quantum Theory*. p. 23. Edited by J. D. Boer, E. Dal, and O. Ulfbeck. Amsterdam: Elsevier.
- Anderson, M. H., J. R. Ensher, M. R. Matthews, C. E. Wieman, and E. A. Cornell. 1995. Observation of Bose-Einstein Condensation in a Dilute Atomic Vapor. *Science* **269**: 198.
- Andrews, M. R., C. G. Townsend, H. J. Miesner, D. S. Durfee, D. M. Kurn, and W. Ketterle. 1997. Observation of Interference between Two Bose Condensates. *Science* **275**: 637.
- Berman, G. P., A. Smerzi, and A. R. Bishop. 2002. Quantum Instability of a Bose-Einstein Condensate with Attractive Interaction. *Phys. Rev. Lett.* **88**: 120402.
- Bradley, C. C., C. A. Sackett, J. J. Tollett, and R. G. Hulet. 1995. Evidence of Bose-Einstein Condensation in an Atomic Gas with Attractive Interactions. *Phys. Rev. Lett.* **75**: 1687.
- Davis, K. B., M.-O. Mewes, M. R. Andrews, N. J. van Druten, D. S. Durfee, D. M. Kurn, and W. Ketterle. 1995. Bose-Einstein Condensation in a Gas of Sodium Atoms. *Phys. Rev. Lett.* **75**: 3969.
- Denschlag, J., J. E. Simsarian, D. L. Feder, C. W. Clark, L. A. Collins, and J. Cubizolles. 2000. Generating Solitons by Phase Engineering of a Bose-Einstein Condensate. *Science* **287**: 97.
- Dziarmaga, J., A. Smerzi, W. H. Zurek, and A. R. Bishop. 2002. Dynamics of Quantum Phase Transition in an Array of Josephson Junctions. *Phys. Rev. Lett.* **88**: 7001.
- Greiner, M., O. Mandel, T. Esslinger, T. W. Hänsch, and I. Bloch. 2002. Quantum Phase Transition from a Superfluid to a Mott Insulator in a Gas of Ultracold Atoms. *Nature* **415**: 39.
- Huang, K. 1987. *Statistical Mechanics*. Second Edition. New York: John Wiley & Sons.
- London, F. 1938. The λ -Phenomenon of Liquid Helium and the Bose-Einstein Degeneracy. *Nature* **141**: 643.
- Madison, K. W., F. Chevy, W. Wohlleben, and J. Dalibard. 2000. Vortex Formation in a Stirred Bose-Einstein Condensate. *Phys. Rev. Lett.* **84**: 806.
- Milonni, P. W., and A. Smith. 1996. van der Waals Dispersion Forces in Electromagnetic Fields. *Phys. Rev. A* **53**: 3484.

- Orzel, C., A. K. Tuchman, M. L. Fenselau, M. Yasuda, and M. A. Kasevitch. 2001. Squeezed States in a Bose-Einstein Condensate. *Science* **291**: 2386.
- Packard, R. E. 1998. The Role of the Josephson-Anderson Equation in Superfluid Helium. *Rev. Mod. Phys.* **70**: 641.
- Pais, A. 1979. Einstein and the Quantum Theory. *Rev. Mod. Phys.* **51**: 861.
- Penrose, O. 1951. On the Quantum Mechanics of Helium II. *Philos. Mag.* **42**: 1373.
- Penrose, O., and L. Onsager. 1956. Bose-Einstein Condensation and Liquid Helium. *Phys. Rev.* **104**: 576.
- Pfleegor, R. L., and L. Mandel. 1967. Interference of Independent Photon Beams. *Phys. Rev.* **159**: 1084.
- Rayfield, G. W., and F. Reif. 1964. Quantized Vortex Rings in Superfluid Helium. *Phys. Rev.* **136**: A1194.
- Sachdev, S. 1999. Chapter 10 in *Quantum Phase Transitions*. Cambridge: Cambridge University Press.
- Timmermans, E. 1998. Phase Separation of Bose-Einstein Condensates. *Phys. Rev. Lett.* **81**: 5718.
- . 2001a. Degenerate Fermion Gas Heating by Hole Creation. *Phys. Rev. Lett.* **87**: 240403.
- . 2001b. Superfluids and Superfluid Mixtures in Atom Traps. *Contemp. Phys.* **42**: 1.
- Timmermans, E., and R. Côté. 1998. Superfluidity in Sympathetic Cooling with Atomic Bose-Einstein Condensates. *Phys. Rev. Lett.* **80**: 3419.
- Timmermans, E., K. Furuya, P. W. Milonni, and A. K. Kerman. 2001. Prospect of Creating a Composite Fermi-Bose Superfluid. *Phys. Lett. A* **285**: 228.
- Timmermans, E., P. Tommasini, R. Côté, M. Hussein, and A. Kerman. 1999. Rarified Liquid Properties of Hybrid Atomic-Molecular Bose-Einstein Condensates. *Phys. Rev. Lett.* **83** (14): 2691.



Eddy Timmermans is currently an Oppenheimer postdoctoral fellow at Los Alamos National Laboratory and will soon join the permanent staff of the Theoretical Division. He received his Ph.D. in condensed-matter theory from Rice University (1995) and then became a fellow at the Harvard-Smithsonian Institute for Theoretical Atomic and Molecular Physics (1995–1998). There, he became interested in the physics of ultracold-atom systems. His most recent research has explored issues of and prospects for fermion pairing in ultracold-atom systems.



Schrödinger Cats

in Atom-Trap BECs

Diego A. R. Dalvit and Jacek Dziarmaga

Microscopic quantum superpositions are routinely observed in experiment. Macroscopic quantum superpositions, on the other hand, are still encountered rarely despite nearly a century of experimentation with quantum mechanics. Fast decoherence of macroscopic states is to blame for this state of affairs (see the articles “Decoherence and the Transition from Quantum to Classical” and “The Emergence of Classical Dynamics in a Quantum World” on pages 86 and 110). And yet, the past few years have witnessed several breakthroughs in the macroscopic regime.

To name a few, superposition states of macroscopic numbers of photons and atoms have been produced in cavity quantum electrodynamics, matter-wave interference in fullerene carbon-60 has been observed, and controlled decoherence due to engineered environments has been measured in ion traps. Recently, the first detection of a macroscopic Schrödinger cat state in a radio-frequency (rf) superconducting quantum interference device, or SQUID (a superposition of clockwise and counterclockwise superconducting current flow), was reported. All these achievements tempt one to try similar investigations of basic quantum mechanics in the rapidly growing field of Bose-Einstein condensates (BECs).

In the article “Atom-Trap BECs” on page 136, Eddy Timmermans describes the possible emergence of nonclassical behavior by number squeezing in a dilute BEC. For a double-well configuration, the ground state of the condensate is determined by the competition

between the tunneling energy $E_{\text{tun}} = -\gamma E_J \cos \alpha$, which favors states with a well-defined relative phase between the wells, and the interaction energy $E_{\text{int}} = (U/2)(N^2 + m^2)$, which favors number states in each well. When the interaction energy is repulsive ($U > 0$), the ground state corresponds to $m = 0$, and $\alpha = 0$, that is, an equal number of particles in the two wells with zero relative phase. However, for attractive interactions ($U < 0$), the ground state is very different: It corresponds to a superposition of states with $m = +N$

$$|\Psi\rangle = \frac{1}{\sqrt{2}} \left[|N_L = N, N_R = 0\rangle + |N_L = 0, N_R = N\rangle \right].$$

and $m = -N$, namely, This state is clearly nonclassical, all N bosons being simultaneously in the left well and in the right well. It corresponds to a macroscopic quantum superposition—a BEC Schrödinger cat—analogue to Schrödinger’s Gedanken experiment of a cat in the weird superposition of being both dead and alive.

Various schemes have been proposed for building macroscopic superpositions in BECs. For example, for a BEC in a double-well potential with an attractive interparticle interaction, one can in principle create the cat state through adiabatically cooling down the BEC until the ground state is reached. Another option is to confine bosons that have an attractive interaction between atoms in two hyperfine levels (A and B) in a single potential well. Initially, all atoms in the BEC are in a given hyperfine state, say A, and then a resonant rf pulse is applied to the sys-

tem to transfer (or rotate) the atoms part of the way between state A and B. The duration of the pulse is much shorter than the self-dynamics of the condensate. At this stage, each atom is in a superposition of levels A and B, and the corresponding many-body quantum state is a product of single-particle superpositions of A and B, that is, it is still a microscopic superposition. However, as this initial state evolves under the nonlinear Hamiltonian that governs the BEC with its attractive interparticle interactions, it reaches a macroscopic superposition in which all atoms are simultaneously in level A and level B, $|\Psi\rangle = (1/\sqrt{2})[|N_A, 0_B\rangle + |0_A, N_B\rangle]$. An even weirder superposition state has been proposed, namely, a coherent superposition of atomic and molecular BECs. It must be stressed that, to date, no experiment has been carried out that attempts to produce any of the aforementioned superposition states.

The condensate is an open quantum system, that is, it is in contact with an environment mainly composed of noncondensed thermal particles. The interaction between that environment and the BEC cat state may cause the loss of coherence between the components of the quantum superposition. If the decoherence time were very small, then the existence of these states in a BEC would be merely of academic interest because there would be no chance of observing them in the laboratory. Therefore, it is important to understand how the thermal cloud affects the longevity of BEC cat states. In principle, a single noncondensed atom colliding with the condensed superposition

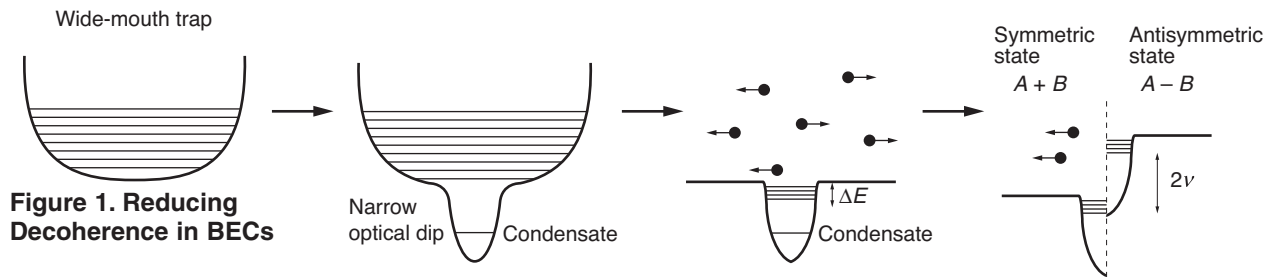


Figure 1. Reducing Decoherence in BECs

state and taking away information about the phase of the state is enough to kill the atomic coherence. Estimated decoherence times for the proposed BEC cat states are inversely proportional to the product of N_E (the number of noncondensed bosons) and N^2 (where N is the number of bosons in the condensate), that is, $t_{\text{dec}} \approx 10^5 \text{ seconds}/(N_E N^2)$. For N_E from 10^0 to 10^4 and N from 10^1 to 10^7 , the decoherence times can range over 16 orders of magnitude, from 1000 seconds down to 10^{-13} second. Given that macroscopic cats require big values of N , it is clear that, for the sake of the cat's longevity, one must go beyond the standard trap settings.

In what follows, we concentrate on a BEC cat formed with two hyperfine states A and B . We show that, by using a combination of trap engineering and what we call "symmetrization" of the environment, as illustrated in Figure 1, one can decrease decoherence rates. First, one prepares the condensate inside a wide magnetic trap and then superimposes a narrow optical dip. The parameters of the traps are chosen such that only a single bound state lies within the dip. The bosons are forced to adiabatically condense into that state. Then the magnetic trap is opened, and most of the noncondensed atoms are allowed to disperse away. The aim of this procedure is to eliminate as much of the thermal cloud as possible. However, atoms occupying bound states within an energy band of width ΔE at the mouth of the dip may not disperse away, but the occupation numbers of those states before the opening of the wide trap may subsist. Those atoms would stay in contact with the conden-

sate and continue to monitor its quantum state and thereby destroy any chance of the condensate to form a superposition. Even if such a truncated environment is relatively harmless, there are ways to better protect the condensate from it.

What we call "symmetrization" of the environmental states can further reduce the decoherence rate. To produce symmetrization, one applies an rf pulse with frequency ν that induces coherent transitions between states A and B of all atoms, both condensed and thermal ones. On the one hand, the state of the condensate is still a macroscopic superposition but slightly different from the original one $(1/\sqrt{2})[|N_A, 0_B\rangle + |0_A, N_B\rangle]$ because the rf pulse produces a small increase in the variance of the number of atoms in each well. On the other hand, the single-particle energy spectrum of the noncondensed bosons is modified. It is now composed of two energy-level ladders shifted with respect to each other by 2ν . One ladder is shifted down, corresponding to states symmetric under the interchange $A \leftrightarrow B$, and the other is shifted upwards, corresponding to states antisymmetric under such interchange. When the energy bandwidth near the mouth of the dip $\Delta E \ll 2\nu$, only symmetric environmental states are occupied. A collision between atoms occupying those states and the condensate does not affect the phase coherence of the latter because both states and the interaction Hamiltonian are symmetric under the interchange $A \leftrightarrow B$. In other words, a symmetric environmental state affects the components $(|N_A, 0_B\rangle$ and $|0_A, N_B\rangle)$ of the BEC cat in exactly

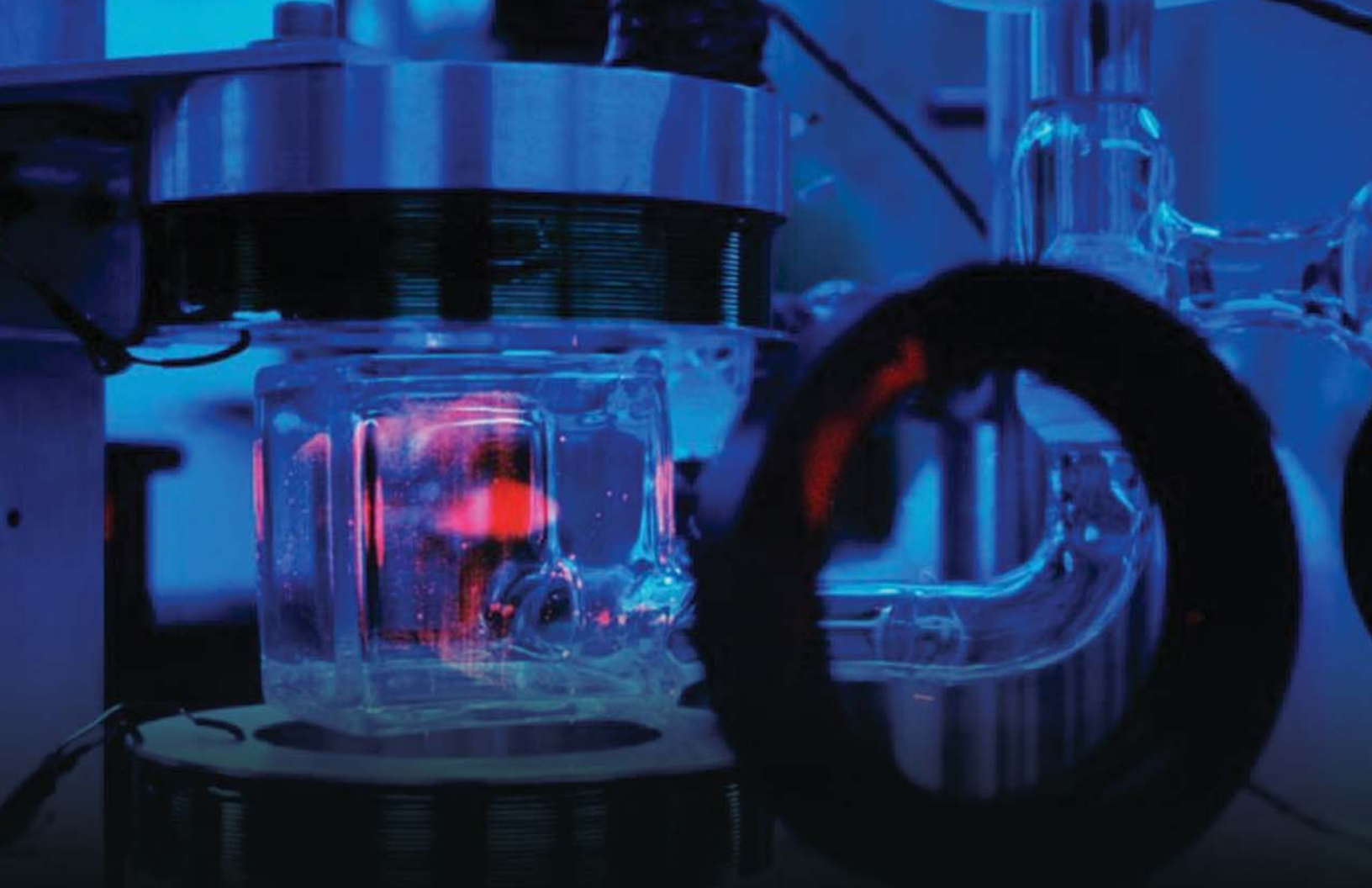
the same way, multiplying them by a common phase factor, which obviously does not affect the phase coherence of the condensate. When the relation $\Delta E \ll 2\nu$ does not hold, some atoms will occupy antisymmetric environmental states and can cause decoherence. However, since that occupation number can be controlled by the intensity of the laser field inducing the coherent transitions between the states A and B , the method of symmetrization can still significantly extend the longevity of the BEC cat. ■

Diego A. R. Dalvit graduated from the Universidad de Buenos Aires, Argentina, in 1993 with a Master's degree in physics and received a Ph.D. in physics from the same university in 1998. Diego has been working as a postdoctoral researcher in the Theoretical Division at Los Alamos since 1999. He is the author of a book on problems in statistical mechanics. He is currently working on the dynamical Casimir effect, decoherence theory, and Bose-Einstein condensation.



Jacek Dziarmaga graduated from the Jagellonian University in Kraków, Poland, in 1994 with a Master's degree in physics and received a Ph.D. in physics from the Jagellonian University in 1995. He has been working as a Senior Research Assistant in the Department of Mathematical Sciences at the University of Durham in the United Kingdom, then as a Director-funded postdoctoral fellow at Los Alamos National Laboratory, and now he is an adjunct professor at the Jagellonian University in Kraków. He is currently working on phase transitions, decoherence theory, and Bose-Einstein condensates.





Experiments on Cold Trapped Atoms

David J. Vieira and Xinxin Zhao

Those of us who have fun trying to take a picture of a fast moving object usually end up with a blurry, imprecise image. Something similar happens when we try to make precision measurements on moving atoms—the movement results in a broadening of intrinsic atomic line widths, and we end up with an imprecise understanding of the subtle atomic processes that produce those lines. Likewise,

detailed studies of the interactions between atoms are hindered by motion because energetic collisions between atoms tend to complicate the system's dynamics and/or mask quantum effects. In general, if we are interested in making precision measurements on the individual or collective properties of free atoms, we have to slow the atoms down.

Kinetic theory tells us that the velocity of an atom in a gas is pro-

portional to the square root of the temperature and inversely proportional to the atom's mass. The atoms and small molecules in the air that we breathe, for example, move about at astonishingly high velocities at room temperature—about 4000 kilometers per hour. Because the velocity varies only as the square root of the temperature, one must make a gas very cold in order to substantially slow the atoms. At one degree above

absolute zero (1 kelvin), atoms still cruise at a few hundred kilometers per hour. Only when temperatures of a few millionths of a kelvin (a few microkelvins) are reached do free atoms move slowly enough that we can make high-precision spectroscopic measurements.

Several methods have been developed that use laser light to cool gases to the microkelvin temperature range. The cold atoms can then be contained within different kinds of atom traps, where they can be studied very accurately or cooled to even lower temperatures. The traps also allow us to concentrate a large number of atoms into a small volume. As the number density increases, the individual atoms begin to “feel” one another, and we can begin to study the transition from individual to collective behavior. With certain “bosonic” atomic species, cooling and trapping techniques enable us to create one of the most fascinating—and fragile—states of matter in the universe, the Bose-Einstein condensate (BEC). See the box “The Bose-Einstein Condensate” on the next page and the article “Atom-Trap BECs” on page 136.

The atom-trapping team at Los Alamos National Laboratory has adapted cooling and trapping techniques to radioactive atoms for both fundamental and applied research. We are in the process of

making sensitive measurements of parity violation in nuclear beta-decay as a means to test the Standard Model of electroweak interactions. We are also trying to cool a dilute gas of fermions to a degenerate quantum state (degenerate Fermi matter), where the density is comparable to that found in a BEC. Aside from displaying interesting quantum mechanical properties, ultracold fermions could undergo a phase transition to a superfluid state, and our apparatus should give us unprecedented control in forming and studying this system. Finally, we are using atom-trapping technology to trap and measure isotopic ratios of selected nuclear species at ultrasensitive levels for nonproliferation treaty verification and environmental studies.

Cooling and Trapping Techniques

Laser cooling of neutral atoms was proposed in 1975 by Theodore Hänsch and Arthur Schawlow, both then at Stanford University. The basic idea was to use the momentum transfer between a photon and an atom to slow the atom down.

When an atom absorbs a photon, its momentum is reduced by an amount $p = h\nu/c$ where h is Planck’s constant, ν is the frequency of the light, and c is the

speed of light. When the atom emits a photon, it gains momentum of the same magnitude (a so-called momentum kick). If, as in laser light, all the absorbed photons come from the same direction, then after many photon scattering events (rapid absorption and emission events), the net change in momentum will be unequal, since the fluorescent photons are emitted in all directions and the sum of the momentum kicks averages to zero. The result is a net loss of momentum.¹

To get laser cooling to work, we use the Doppler effect to ensure that only those atoms moving into the laser beam will absorb photons. The Doppler effect relates the intrinsic frequency of a source to the frequency “sensed” by an observer moving relative to the source. The pitch of a siren, for example, sounds higher when we move quickly toward it (or it moves quickly toward us) and lower when we move rapidly away. Similarly, an atom “sees” the frequency of a photon increase when the atom moves toward the photon. Thus, if we tune a laser to have a

¹ The change in momentum due to light scattering means that the atom feels a pressure, which can be quite large (up to 10,000 times larger than the force of gravity). Radiation pressure provides a very effective means of moving atoms around.

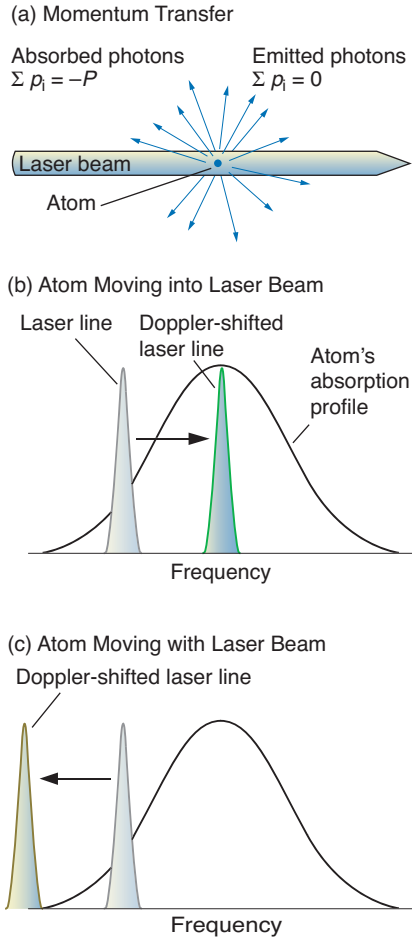


Figure 1. Laser Cooling

(a) An atom illuminated by laser light will absorb and reemit (scatter) many photons. (b) If the laser frequency is tuned below the atomic resonance line (red detuned), then an atom moving against the laser beam “sees” a laser frequency that is Doppler-shifted closer to the absorption maximum. It absorbs the low-energy laser photons. The atom then emits a higher-energy photon at the resonance frequency of its transition line. The atom loses energy with each absorption/emission event and begins to cool. (c) An atom moving in the same direction as the laser beam “sees” the detuned laser frequency Doppler-shifted still farther away from its absorption maximum. The atom absorbs few photons and is not cooled.

The Bose-Einstein Condensate

Elementary particles—and collections of particles such as nuclei and atoms—are either fermions (and have half integer spin) or bosons (and have integer spin). In the mid-1920s Albert Einstein, building on the work of Satyendra Nath Bose, predicted that, at exceptionally low energies, an ensemble of massive bosons should undergo a transition into a state that is described by a single, coherent wave function. This coherent state—now called the Bose-Einstein condensate (BEC)—would be as different from ordinary matter as laser light is from sunlight.

Physicists believed that a dilute gas of bosons could form a BEC, but the conditions needed to produce one are extreme. In order to become coherent, or establish a common phase relationship amongst themselves, the atomic wave functions must overlap significantly with one another. The spatial extent of the atomic wave function is given by its de Broglie wavelength λ , and it can be shown that the BEC will form if the atom density, expressed as the number of atoms in a λ -sided cube, exceeds 2.6. Both the de Broglie wavelength and the density of a gas depend on temperature, and one can calculate how cold it must be to achieve the critical density in a cold boson gas. The answer is, on the order of a few hundred billionths of a kelvin.

Certainly, one problem in creating a BEC was to find a gaseous system that would not coalesce into a solid as the temperature plunged toward absolute zero. The solution was to use certain alkali atoms (atoms from group I of the Periodic Table). When spin-polarized, these atoms have a weak repulsive force between them that would ensure that the system remained a gas. A BEC of rubidium-87 atoms was finally created and observed in 1995 by Carl Weiman’s and Eric Cornell’s group at the University of Colorado / JILA (Joint Institute for Laboratory Astrophysics). Four months later, Wolfgang Ketterle’s group from the Massachusetts Institute of Technology created a BEC from sodium-23 atoms. Since that time, a BEC has been observed in several other bosonic alkali species, such as hydrogen-1 and lithium-7. All the efforts involved cooling the atoms (except hydrogen atoms) to less than a millikelvin in what is called a magneto-optical trap (MOT), reducing the temperature by another order of magnitude by laser cooling, and then transferring the atoms to a magnetic trap. There, the atoms are cooled by a technique known as evaporative cooling to less than 200 nanokelvins to create a BEC.

slightly lower frequency than the resonance frequency of an atom’s

absorption line (detuning), only atoms that happen to be moving against the beam see the frequency of the photon Doppler-shift into resonance (see Figure 1). These atoms lose momentum and are slowed down (cooled). Atoms moving in the same direction as the detuned laser beam are Doppler-

shifted farther away from resonance. They do not readily absorb photons and are consequently unaffected.

To cool the atoms in three dimensions requires six intersecting laser beams—one pointing in each of the six directions $\pm x$, $\pm y$, and $\pm z$. Then any atom that emerges from the intersection region will be moving against a properly tuned laser beam and will be cooled.

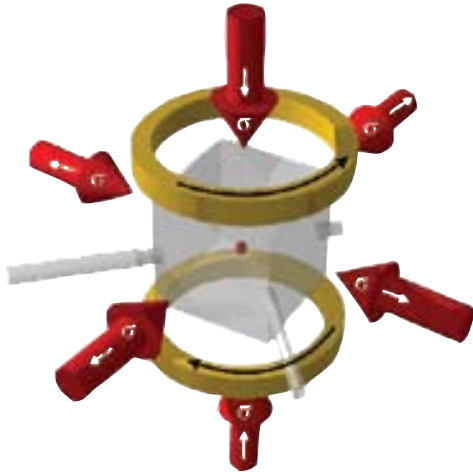


Figure 2. The Magneto-optical Trap (MOT)

(a) The MOT consists of six circularly polarized laser beams that intersect at the zero point of a magnetic field (produced by the set of anti-Helmholtz magnetic coils). The tube projecting from the left is used to bring atoms into the evacuated glass cell located between the coils. (b) This schematic energy diagram indicates why trapping occurs.

The σ^- polarized light induces a transition from the ground state $|S, m_s\rangle = |0, 0\rangle$ to the $|1, -1\rangle$ excited state, whereas the σ^+ polarized light will induce a transition from $|0, 0\rangle$ to $|1, +1\rangle$. The atom's magnetic substates are Zeeman-split by the magnetic

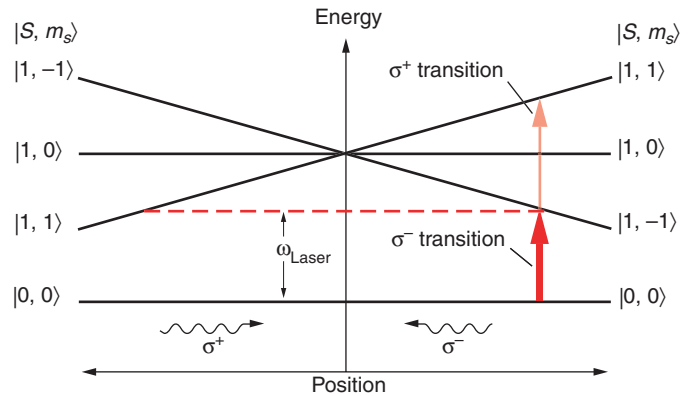
field. The force experienced by an atom during laser cooling is velocity dependent; that is, its magnitude depends on the atom's velocity as it moves toward the laser beam. (The three-dimensional laser cooling is often called an optical molasses because velocity-dependent forces are viscous forces and the atom behaves as if it were entrained in a viscous liquid. The term optical molasses was coined by Steven Chu of Stanford University.) Velocity dependence means that the cooling rate decreases as the atom slows down. When the velocity gained by the atom as it emits a photon (the atom recoil) equals the loss of velocity due to the scattering process, the cooling ceases altogether. The minimum velocity of the atom at the "recoil limit" translates into a mini-

² There are also subrecoil laser-cooling mechanisms that can cool atoms below the recoil limit.

imum temperature.² For sodium atoms, the recoil limit is 2.4 microkelvins and for somewhat heavier cesium atoms it is about 0.2 microkelvin.

The Magneto-optical Trap (MOT). Although optical molasses cools atoms down to very low temperatures, the atoms can diffuse out of the laser region through random Brownian motion. The MOT was invented to prevent this loss and to confine the atoms. The idea behind the MOT is to combine the optical molasses with an external magnetic field and thereby create a spatially dependent force that acts only on atoms that wander from the trap's center. The MOT was fully developed in David Pritchard's laboratory at MIT in 1987. Because of its relative ease of construction and great utility, it is perhaps the most commonly used atom trap.

For this trap, three pairs of counterpropagating, circularly polarized



field. As the atom drifts away from the center of the MOT, say, to the right of the diagram, an atomic transition to the $m_s = -1$ substate shifts onto resonance with the σ^- polarized laser and starts to preferentially absorb these photons over the σ^+ polarized laser coming from the opposite direction. The resulting laser-induced pressure "pushes" the atom back toward the center. The result is the same if the atom moves out in any direction from the center of the trap.

[Part (b) of the figure was adapted from *Phys. Rev. Lett.* 59 (1987), p. 2631, with permission from the authors.]

laser beams (σ^+ and σ^- polarizations) establish an optical molasses within a vacuum chamber, as seen in Figure 2. Outside the molasses region are two magnetic coils. The current in each coil runs in opposite directions (anti-Helmholtz configuration) and creates a "quadrupole" magnetic field, which has zero field value at the center and the field gradient increases linearly as one moves out from the center in any direction.

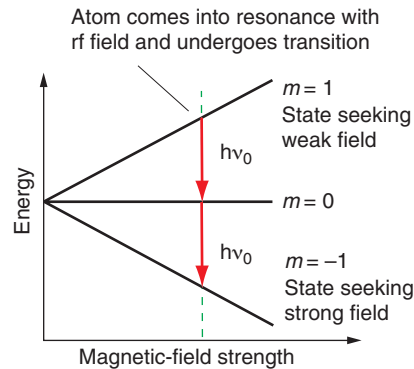
The trap works because an atom's magnetic substates (m -states) have different energies in a magnetic field (the Zeeman effect), and due to the field gradient, the m -state energy increases (or decreases) as the atom moves out from the center of the MOT. With reference to Figure 2(b), an atom in the trap will be illuminated with both σ^+ and σ^- circularly polarized laser light. Suppose the atom moves away from the center of the trap, say, in the $(+z)$ -direction, so that it moves into the σ^- laser beam, but

in the same direction as the σ^+ laser beam. Both lasers are tuned slightly below the $|S = 0\rangle \rightarrow |S = 1\rangle$ resonance frequency. At some distance from the MOT center, the drifting atom will come into resonance with the incoming σ^- radiation (but not with the σ^+ light). Similar to the way in which it absorbs light in an optical molasses, the atom will begin to absorb more of the σ^- light and will feel a pressure that pushes it back toward the center of the MOT. Likewise, an atom moving in the ($-z$)-direction (or $\pm x, \pm y$ directions) will preferentially absorb photons from the inward-directed laser beam and will be pushed back toward the trap's center. Because the magnetic field is symmetric, the atom becomes trapped in three dimensions.

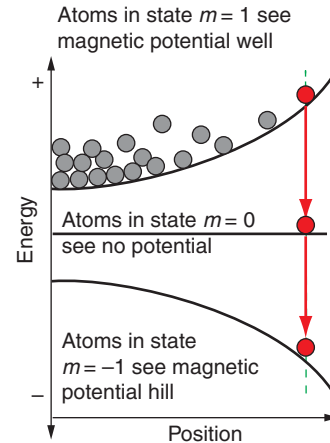
Magnetic Traps, Evaporative Cooling, and the Time-Orbiting Potential (TOP). While the MOT requires lasers to trap the atoms, magnetic fields alone can create a trapping potential. A pure magnetic trap makes use of the fact that atoms will experience a magnetic dipole force in a magnetic field gradient $\mathbf{F} = -\boldsymbol{\mu} \cdot \nabla \mathbf{B}$, where $\boldsymbol{\mu}$ is the atom's magnetic moment and $\nabla \mathbf{B}$ is the magnetic field gradient. If the atom is polarized into the $|m = 1\rangle$ substate, the force will be toward lower magnetic-field values. The atom is diamagnetic and can be trapped by a simple magnetic quadrupole field, which has a zero magnetic-field value at the center.

Magnetic traps are easy to construct, but they have fairly weak trapping potentials (about 1000 times weaker than found in a MOT). They can only trap atoms that are already very cold, with thermal energies equivalent to 1 millikelvin or less. Once inside a magnetic trap, the atoms can be cooled to the limits of laser cooling. To reach the temperatures needed to create a BEC, however, we need another cooling technique, namely, evaporative cooling.

(a) Zeeman-Split Magnetic Sublevels



(b) Principles Underlying Evaporative Cooling


Figure 3. Evaporative Cooling

(a) The figure shows the magnetic sublevels of an atom as a function of magnetic-field strength. An atom in the state $|m = 1\rangle$ is diamagnetic because it has lower energy in weaker magnetic fields. (Therefore the atom is attracted to regions of weaker field.) Conversely, atoms in the state $|m = -1\rangle$ are paramagnetic (that is, attracted to regions of higher magnetic fields). If the atom is illuminated by an rf radiation of frequency ν_0 , then at some magnetic-field value, the atom can come into resonance with the radiation and undergo a transition from $|m = 1\rangle$ to $|m = 0\rangle$, and then from $|m = 0\rangle$ to $|m = -1\rangle$. The diamagnetic atom converts into a paramagnetic one. (b) The evaporative cooling technique removes the most energetic atoms from a magnetic trap. Atoms in the trap are polarized in the $|m = 1\rangle$ (diamagnetic) state and are trapped by the quadrupole magnetic field. The most energetic atoms make the greatest excursions from the trap center and move into regions of higher magnetic field. These atoms come into resonance with an rf field and are converted to paramagnetic atoms, which are ejected from the trap. (They move to high-field regions outside the trapping volume.) After reequilibration through atomic collisions, the remaining atoms reach a lower temperature.

Temperature is a measure of the average kinetic energy of a system, and in a gas, the energy is distributed amongst the atoms according to a Maxwell-Boltzmann distribution. This means that some atoms always have greater than the average energy. We can efficiently cool a gas by removing the highest-energy atoms. After the remaining gas re-equilibrates, it will have a lower average energy. The common name for this process is evaporation. A liquid that is evaporating (say a steaming cup of coffee) cools down because the most energetic atoms leave (and form the rising steam).

To further cool the already cold

atoms, we actively eject the most energetic particles. We stated that the magnetic trap holds onto diamagnetic atoms. But atoms polarized in the $|m = -1\rangle$ substate are paramagnetic and will be attracted to the higher magnetic fields outside the trapping region. A radio-frequency (rf) field can be used to induce transitions between magnetic substates and convert an atom that is diamagnetic to one that is paramagnetic, at which point it is ejected from the trap. The frequency of the rf field is chosen such that only atoms with enough energy to move to the edge of the magnetic potential well come into resonance with the rf field (see Figure 3).

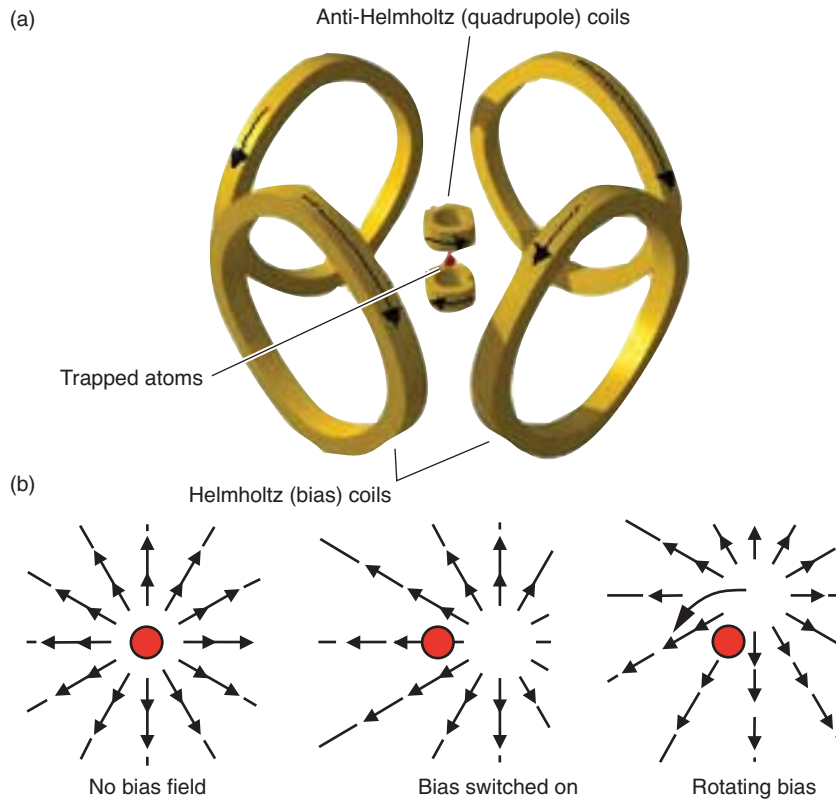


Figure 4. The Time-Orbiting-Potential (TOP) Trap

(a) The TOP trap is a magnetic trap that combines two magnetic fields: a quadrupole field (produced by the central, anti-Helmholtz coils) and a bias field (produced by the outer Helmholtz coils). (b) With the addition of the bias field, the potential minimum of the magnetic trap shifts off-axis. By adjusting the current in the Helmholtz coils, we make the bias field rotate around the trap axis and produce a time-averaged total field with a positive field strength at the center of the trap. As long as the bias field rotates fast enough, the atoms will remain polarized and stay trapped.

After ejecting the most energetic atoms from the trap, the rf frequency is readjusted so that once again the most energetic atoms of the now colder gas are ejected. In this way, it is possible to successively skim off the hottest atoms and thereby evaporatively cool the atoms.

One problem with this cooling scheme is that the quadrupole field has zero field strength at the center of the trap. Consequently, the magnetic substates are not Zeeman-split at the center of the trap, so polarized atoms can undergo spontaneous spin-flip transitions to the $|m = 0\rangle$ or $|m = -1\rangle$ substates in this region. The loss rate by this mechanism increases as the atoms become colder, making it difficult to achieve the critical BEC conditions of high atom density and low temperature.

The TOP trap, developed by Eric Cornell and collaborators, eliminates this problem by adding an off-axis bias field to the static quadrupole

field. As seen in Figure 4, the minimum of the total magnetic field becomes shifted away from the trap center. By rotating the bias field, the time-averaged total field still retains its basic quadrupole configuration, but now it has positive field strength at the center, so the atoms remain polarized. The bias field must rotate faster than the atoms can respond,³ but this objective is easily achieved. The TOP trap allows the density of atoms in the trap to increase sufficiently as the atoms are evaporatively cooled to reach the conditions for a BEC.

Atom Trapping at Los Alamos

³ The atoms oscillate within the harmonic potential well of the TOP trap. If the atoms are to experience the time-averaged magnetic field, the bias field must rotate faster than the atoms' period of oscillation.

Having cold, almost frozen, atoms at our disposal allows us to perform high-precision experiments to test quantum theories of ultracold ensembles of atoms and the nature of fundamental forces. Our system at Los Alamos, illustrated in Figure 5, combines several of the techniques and traps discussed above. A high-efficiency MOT that is coupled to an off-line mass separator is used for trapping radioactive atoms. Once the atoms are trapped, they can be counted with high sensitivity (via fluorescence detection) or transferred to another trap, where various experiments can be performed. At present, we are pursuing a number of research initiatives.

Parity Violation in Nuclear Beta-Decay. Spatial reflection symmetry, otherwise known as parity conservation, maintains that the fundamental processes of nature should be the same under a spatial inversion of all

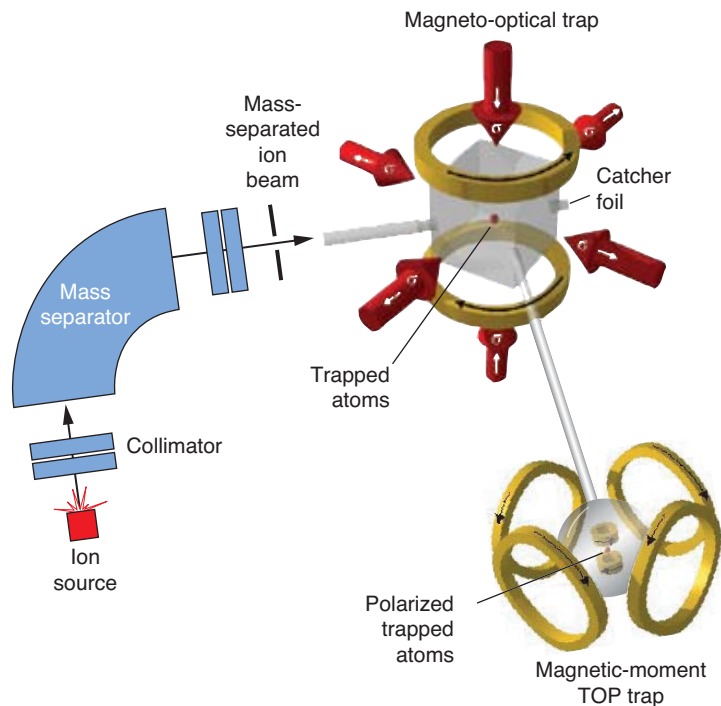


Figure 5. Los Alamos Setup to Trap Radioactive Atoms

Cooling atoms to ultralow temperatures must be done in stages, with several traps and laser configurations. In the setup at Los Alamos, energetic, radioactive atoms from an ion source are implanted into a thin metal foil that sits within an evacuated glass cell, around which are the MOT field coils. Heating the foil releases the atoms into the cell where they are trapped in the MOT and cooled to about 100 μK . The MOT is turned off, and a laser pushes the atomic cloud into the evacuated chamber of a second MOT, where the atoms are recaptured. The magnetic field of this second MOT is turned off and an optical molasses is established by detuning the frequency of the laser further to the red (that is, to lower frequency). Within a few milliseconds, the atoms have cooled to approximately 20 μK , and then they are optically pumped into a diamagnetic substrate with a polarized laser beam. The optical pumping beam is then turned off, and the magnetic field is quickly ramped up in a TOP configuration. We plan to use evaporative cooling to bring the atoms to a final temperature of a few hundred nanokelvins.

vector parameters. Parity conservation was verified in electromagnetic and strong interactions, but as a startled physics community discovered in the 1950s, not in the weak interaction. Despite the astounding progress that has been made in understanding fundamental forces over the past fifty years, the origin of parity violation in the weak interaction remains a mystery of modern science. We hope to make a very precise measurement of the degree of parity violation in rubidium-82 as a means to test current theories.

One way the weak interaction man-

ifests itself is through a type of nuclear beta-decay, whereby a proton in a parent nucleus decays to a neutron, a positron (also known as a beta particle) and an electron neutrino. A daughter nucleus with a different atomic number is created in the process. For example, in rubidium-82 decay,



For the initial and final states of interest, this decay involves pure Gamov-Teller transitions that proceed solely

through the axial-vector (parity-violating) component of the weak interaction and is predicted by the Standard Model to be maximally parity violating. If the rubidium-82 nucleus is polarized by a magnetic field, then parity violation would manifest itself as an asymmetry in the angular distribution of the emitted positrons relative to the nuclear spin direction. For the primary beta-decay branch (in which the rubidium-82 nucleus decays to the 0^+ ground state of krypton-82), the positron is emitted in the direction of the nuclear spin. (In a secondary, less probable decay branch, the positron comes out in a direction opposite to that of the nuclear spin.)

We have recently demonstrated the trapping of polarized, radioactive rubidium-82 atoms. A radiochemically separated sample of strontium-82 ($t_{1/2} = 25$ days) is loaded into the ion source of a mass separator. The strontium-82 decays by electron capture to rubidium-82 ($t_{1/2} = 76$ seconds). The rubidium-82 atoms are thermally ionized, electrostatically extracted, mass separated, and implanted into a zirconium catcher foil located inside a glass cell that sits at the center of a high-efficiency MOT. Heating the foil releases the atoms as a dilute vapor into the glass cell where they are trapped and cooled.

The atoms are rapidly transferred to a second chamber by resonant laser light “pushing” on them. In the second chamber, the atoms are retrapped in a second MOT, further cooled, optically pumped into a specific magnetic substate, and loaded into a TOP magnetic trap. Being in a stretched state, the nuclear spin is aligned with the overall spin of the atom. Consequently, the nuclei are polarized and aligned with the local field. In the center of the TOP, the strongest field is in the direction of the bias field, so the direction of the nuclear spin rotates with the bias field.

By keeping track of the varying

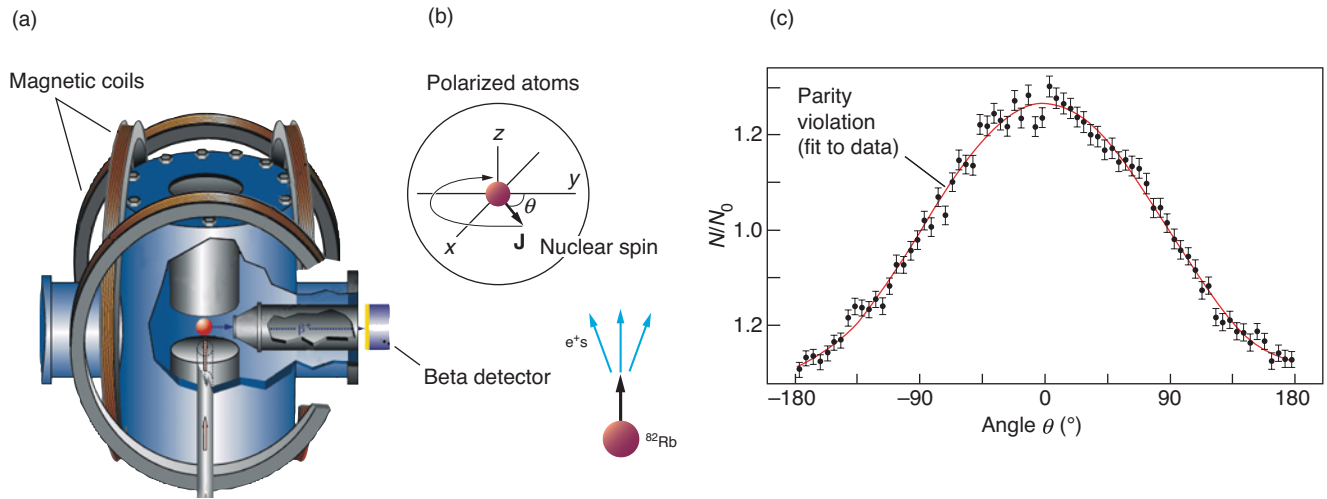


Figure 6. Measurement of Parity Violation

(a) The figure shows the TOP trap of our experimental system. The nuclei of the trapped rubidium-82 atoms are spin polarized and always point in the direction of the TOP's rotating bias field. By monitoring the currents that produce the bias field at any given time, we can reconstruct the magnetic-field orientation; hence, we know the nuclear spin direction. A plastic scintillator is used to detect the emitted positrons. When a positron is detected, we reconstruct the nuclear-spin direction and can

determine the angle θ between it and the positron emission direction. (b) Because parity is not conserved in the weak interaction, the spin-polarized rubidium-82 nuclei will decay by preferentially emitting positrons in the direction of the nuclear spin. (c) This plot of rubidium-82 beta-decay data, accumulated over a period of 6 hours for positrons with energies above 800 keV, shows the parity violating the angular distribution of the positrons. The red line is a cosine fit to the distribution.

currents in the bias coils of the TOP trap, we can reconstruct the direction of the bias field, hence the spin alignment, as a function of time. We can then correlate each beta event with the orientation of the nuclear spin, and record the angle between the beta and the nuclear-spin direction. In Figure 6, we show our initial proof-of-principle results, which indicate that parity is, as expected, violated in the beta decay of polarized rubidium-82 atoms. This is the first time that the entire angle-dependent parity-violating amplitude has been measured.

We are in the process of making a 1 percent measurement of this correlation in order to place stringent limits on the maximal parity-violating nature of the weak interaction. We hope to extend that measurement to 0.1 percent and to search for new physics beyond the Standard Model.⁴

⁴ This work is done in collaboration with scientists from the Chemistry and Physics Divisions at Los Alamos. See Crane et al. 2001.

Ultracold Atoms / Quantum Degenerate Matter. The ability to trap and cool different isotopes enables us to explore mixed fermionic and bosonic systems. In particular, we are working to produce a BEC of bosonic rubidium-87 and overlap it with a magnetically trapped cloud of radioactive, fermionic rubidium-84. In doing so, we hope to sympathetically cool, via atomic collisions, the rubidium-84 atoms down to the Fermi degenerate regime (approximately 10 to 100 nanokelvins). We want to study the fermion-fermion and fermion-boson collision dynamics at temperatures approaching absolute zero.

Recent calculations show that rubidium-84 is a good fermionic candidate for sympathetic cooling because it has a large and positive scattering length with rubidium-87. Calculations also indicate, however, that, in the presence of a relatively low magnetic field ($B \sim 100$ gauss), a Feshbach resonance should be present in rubidium-84. This resonance allows two colliding atoms to form a temporary molecule before separating, and

by adjusting the magnetic-field value, we can fine-tune the energy at which the resonance occurs. In doing so, we can control the collision cross section and effectively “tune” the temperature at which a phase transition to a superfluid state will occur.

The radioactive rubidium-84 atoms ($t_{1/2} = 33$ days) for our experiments are produced by proton spallation reactions on a molybdenum target at the Los Alamos Neutron Scattering Center. The rubidium is chemically extracted from the molybdenum and loaded into the ion source of a mass separator. The rubidium-84 is implanted and captured in the MOT in a similar procedure to that described in the previous section.

As an initial step toward achieving our goal, we have demonstrated the trapping of rubidium-84. Figure 7(a) shows the time-dependent trapping signal from roughly one million trapped rubidium-84 atoms. At high atom densities, the losses from the trap are dominated by laser-light-assisted collisions between trapped atoms.

By overlapping a cloud of 3×10^5

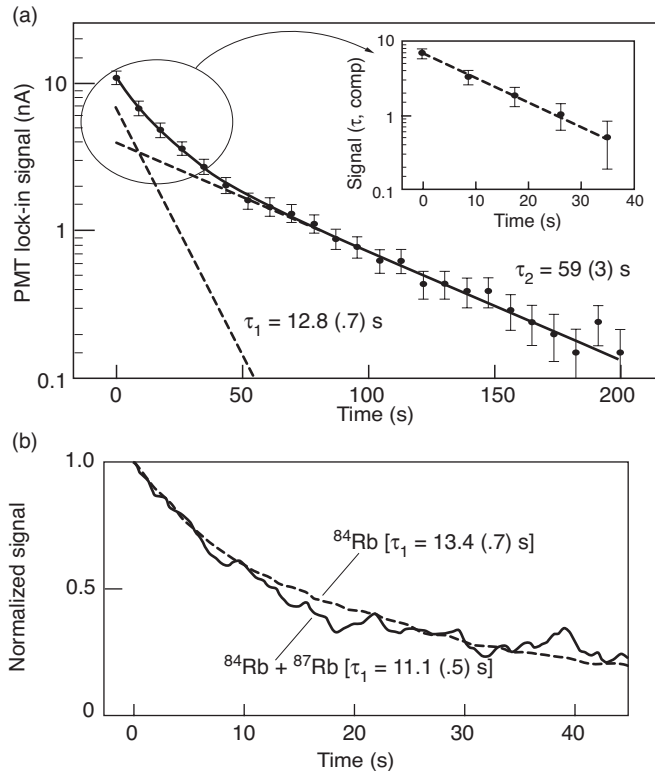


Figure 7. Lifetimes in a MOT: Rubidium-84 with and without an Overlapping Cloud of Rubidium-87

How long will a MOT confine half a million rubidium-84 atoms? The data indicates that the answer depends on the atomic density. (a) At high densities (short times), light-assisted collisions between trapped atoms dominate. These give rise to the short-lived component ($t_1 = 12.8$ s) of the overall trap lifetime. (The inset shows a fit to the short-lived component.) As the number of trapped atoms decreases and the density goes down, light-assisted losses become negligible and only collisional losses between the cold atoms and the hot background gas remain. The collisional losses correspond to the long-lived component, with a lifetime of about 59 seconds. (b) Introducing rubidium-87 atoms into the trap could lead to collisions between the rubidium isotopes and an enhanced loss rate. This figure shows the normalized lifetime in the trap of rubidium-84 atoms with and without an overlapping cloud of rubidium-87 atoms (solid line and dashed line, respectively). The additional loss rate is sufficiently small that it does not present a problem for the sympathetic cooling experiment discussed in the text.

cold atoms of rubidium-84 with a large cloud of 7×10^7 atoms of stable rubidium-87, we have also been able to set a limit on the inelastic-collision loss rate of the atoms from the trap, which could affect the rubidium-84 trapping lifetime—see Figure 7(b). Fortunately, this loss rate was found to be sufficiently small and did not present a problem for the sympathetic cooling experiment. We are currently optimizing the evaporative-cooling

process to achieve quantum degeneracy for the bosonic rubidium-87 and to study its cooling of rubidium-84 (Crane et al. 2000).

Ultrasensitive Detection. As a result of fallout from atmospheric nuclear tests, the two radioactive isotopes cesium-135 ($t_{1/2} = 2.3 \times 10^6$ years) and cesium-137 ($t_{1/2} = 30$ years) are ubiquitous in the environment, at a relative abundance of roughly 1 part per billion with respect to stable cesium-

133. (The fission product isotopes are man-made, that is, anthropogenic.) Cesium adsorbs strongly and rapidly to soil particles, and because the heavier isotope cesium-137 is relatively easy to detect through gamma-ray spectrometry, it has served as a chronometer and tracer in a diverse array of scientific endeavors, including studies of soil erosion and lake sedimentation.

The long radioactive lifetime of cesium-135, however, severely limits its detection by gamma-ray spectrometry. This is unfortunate, since a measurement of the cesium-137/cesium-135 isotope ratio would lead to a relatively unambiguous determination of a sample's age. Furthermore, that particular ratio is of interest for nonproliferation and treaty verification because the cesium-137/cesium-135 content of nuclear-fuel effluent can provide valuable information about nuclear-reactor operations.

Detecting both isotopes, especially from random environmental samples, requires that we have a highly sensitive, highly selective technique. Several advanced technologies, including resonant ionization mass spectrometry (RIMS), have been successfully applied to the problem, with the RIMS method achieving a detection limit of about 1×10^8 atoms, an estimated isotopic selectivity of about 10^{10} , and an overall efficiency (from source size to detectable sample size) of 2×10^{-6} .

We recognized that, when coupled to a mass separator, a MOT could do even better. Because the trapping potential of a MOT derives from a multiphoton, near-resonant absorption process, it is very species selective (atomic, isotopic, and isomeric) with respect to what it traps. The mass separator also has high isotopic selectivity, so a mass separator/MOT system affords a huge suppression of signals from unwanted species. A MOT “detector” should also have high sen-

sitivity. Each trapped atom can scatter (rapidly absorb and emit) about 10^7 photons per second, so even small numbers of atoms can be detected.

We are the first group to have succeeded in trapping and detecting cesium-135 and cesium-137 in a MOT. A sample containing both isotopes was placed in the source of a mass separator, and each isotope was sequentially measured with a MOT. Trapped-atom numbers in the case of either isotope ranged from 10^4 to 10^7 , as determined from the MOT fluorescence signal. Over this trapped-atom range, the MOT fluorescence signal was found to increase linearly with the number of atoms implanted into the foil with no sign of an isotopic dependence to within 4 percent.

Direct measurement of the cesium fluorescence signals should yield the cesium-137/cesium-135 ratio. In principle, our mass separator/MOT technique can make that determination to within 10 percent of uncertainty. Currently, the system has a detection limit of about 10^6 atoms, an isotopic selectivity of greater than 10^{12} , and an overall efficiency of 0.5 percent. As such, our work represents a significant advance in efficiency and isotopic selectivity among other methods applied to the detection of cesium radioisotopes (Di Rosa et al. 2002.). More important, our results demonstrate the advantages of applying atom-trapping techniques to the general problem of ultrasensitive detection.

Conclusions

Over the last decade, advances in the laser cooling and trapping of atoms have revolutionized the prospects of fundamental research and applied quantum-based projects. In atomic physics, scientists have gained unprecedented control over quantum ensembles, as witnessed by the cre-

ation and wide study of BECs today. But the new trapping and cooling techniques should not be viewed as simply a workhorse for quantum optics and atomic physics. Their use has spread to nuclear physics (as in our rubidium-82 experiment), biophysics, condensed-matter physics, quantum information, and environmental science (as demonstrated by our cesium experiments). The results of this “cross-fertilization” have in turn enriched the field of atomic physics. We believe the atom-trapping revolution is just beginning and that in the years to come there will be many new exciting interdisciplinary opportunities. ■

Further Reading

- Crane, S. G., X. Zhao, W. Taylor, and D. J. Vieira. 2000. Trapping an Isotopic Mixture of Fermionic ^{84}Rb and Bosonic ^{87}Rb Atoms. *Phys. Rev. A* **62**: 011402(R).
- Crane, S. G., S. J. Brice, A. Goldschmidt, R. Guckert, A. Hime, J.J. Kitten, D. J. Vieira, and X. Zhao. 2001. Parity Violation Observed in the Beta Decay of Magnetically Trapped ^{82}Rb Atoms. *Phys. Rev. Lett.* **86**: 2967.
- Di Rosa, M. D., S. G. Crane, J. J. Kitten, W. A. Taylor, D. J. Vieira, X. Zhao. 2002. Magneto-Optical Trap and Mass Separator System for the Ultrasensitive Detection of ^{135}Cs and ^{137}Cs , Los Alamos National Laboratory document, LA-UR-02-3926. (Submitted to *Appl. Phys. B*).
- Inguscio, M., S. Stringari, and C. E. Wieman, eds. 1999. *Bose-Einstein Condensation in Atomic Gases, Proceedings of the International School of Physics Enrico Fermi*. Amsterdam: IOS Press.
- DeMarco, B. 2001. “Quantum Behavior of an Atomic Fermi Gas.” Ph.D. thesis, University of Colorado.
- Metcalf, H. and R. van der Straten. 1994. Cooling and Trapping of Neutral Atoms. *Physics Reports* **244**: 203.
- Greiner, M., O. Mandel, T. Esslinger, T. W. Hänsch, and I. Bloch. 2002. Quantum Phase Transition from a Superfluid to a Mott Insulator in a Gas of Ultracold Atoms. *Nature* **415**: 39.
- Roberts, J. L. 2001. “Bose-Einstein Condensate with Tunable Atom-Atom Interactions: The First Experiments with ^{85}Rb BECs.” Ph.D. thesis, University of Colorado.

David J. Vieira is the nuclear chemistry team leader in the Isotope and Nuclear Chemistry Group at Los Alamos. He obtained his Ph.D. in nuclear chemistry from the University of California, Berkeley, in 1978, after which he came to Los Alamos. He has been an Adjunct Professor of Physics at the Utah State University since 1984 and was an Alexander von Humboldt Fellow (1990–91). David received a Los Alamos Distinguished Team Performance Award in 1993 for the development of a He-jet system at the Los Alamos Neutron Science Center (LANSCE) and the Fellows Prize (1998) for the development of the time-of-flight isochronous (TOFI) recoil spectrometer, the study of exotic nuclei using TOFI, and the trapping of radioactive atoms. David’s current interests include fundamental atomic and nuclear physics experiments involving trapped radioactive atoms, ultrasensitive detection, quantum information and control using trapped atoms, as well as fundamental symmetries, radioactive beams, and neutron-induced cross section measurements.



Xinxin Zhao received his Ph. D. from Rice University in 1993. After three years of post-doctoral research on single-ion high-precision spectroscopy at the University of Washington, he came to Los Alamos National Laboratory to work on laser cooling and trapping of radioactive atoms. In 1999, he became a technical staff member at the Laboratory. He has over ten years of experience in the areas of atom/ion trapping, laser cooling, and high-resolution laser spectroscopy. Together with David Vieira and collaborators, Xinxin has demonstrated the trapping of a record number of radioactive atoms and observed the first beta-decay asymmetry spectrum from magnetically trapped polarized atoms. His current research interests are application of laser cooling and trapping to fundamental and applied physics.





Quantum Information with Trapped Strontium Ions

Dana J. Berkeland

Something wonderful happens when small numbers of ions are trapped in a linear Paul (radio-frequency, or rf) trap and laser-cooled. The ions become nearly motionless and line up neatly along the trap axis—each confined to its own tiny space of about 100 micrometers or less in any direction. Because the ions are frozen in place, experimental physicists can continually observe them for up to months at a time and gain uncommon insight into the quantum realm.

For example, single ions exhibit quantum-mechanical effects that could never be observed in a large ensemble of ions or neutral atoms. A large field of study in quantum optics has in fact emerged with the development of ion traps (Thompson et al. 1997). In addition, the internal transitions of a nearly motionless ion are only slightly affected by Doppler shifts, and the ion can be superbly isolated from unwanted electric fields and noisy magnetic fields. This characteristic makes a trapped ion a useful testing ground for many physical theories that

predict very small shifts of the atomic energy levels (Berkeland et al. 1999). Finally, a focused laser beam can interact first with one specific ion, then a different one—a capability that means we can control complicated interactions between states of a particular ion and between different ions. For this reason, the ion trap has shown considerable promise as the basis for a quantum computer. (See the article “Ion-Trap Quantum Computation” on page 264.)

In this article, I discuss some of our activities with trapped and laser-cooled ions. I focus on an experiment that provides a fundamental test of quantum-mechanical randomness but also mention a spectroscopy experiment that is a prerequisite to the development of a quantum logic gate. For background material, see the previously mentioned article, “Ion-Trap Quantum Computation,” which discusses the operational principles of a linear Paul trap and laser cooling.

We conduct our experiments using singly ionized strontium atoms. Figure 1(a) is an illustration of our linear Paul trap (Berkeland 2002).

Most of the trap has been created with off-the-shelf components and requires no precise or otherwise demanding machining to assemble. This feature is significant because it shows that ion trapping with linear traps can be an accessible technology for groups with limited resources.

Figure 1(b) shows the transitions we use in the strontium ion $^{88}\text{Sr}^+$. We use the 422-nanometer transition to Doppler-cool the ions. We also collect the 422-nanometer fluorescent light from the decay of the $P_{1/2}$ state and focus it onto a detector to image the ions. Light at 1092 nanometers drives the $D_{3/2} \leftrightarrow P_{1/2}$ transition to prevent the atoms from pooling in the long-lived $D_{3/2}$ state, in which they would not scatter any 422-nanometer light. A 674-nanometer diode laser drives transitions between the $S_{1/2}$ ground state and the $D_{5/2}$ state, which lives an average of 0.35 seconds. This transition can be used to couple the $S_{1/2}$ and $D_{5/2}$ states of the ion with its motional states, any of which may be used as qubits in a quantum computer. The $S_{1/2} \leftrightarrow D_{5/2}$ transition is also driven

About forty strontium ions lined up in our linear Paul trap are visible because they scatter laser light. The apparent gaps are due to other ions that do not scatter the light.

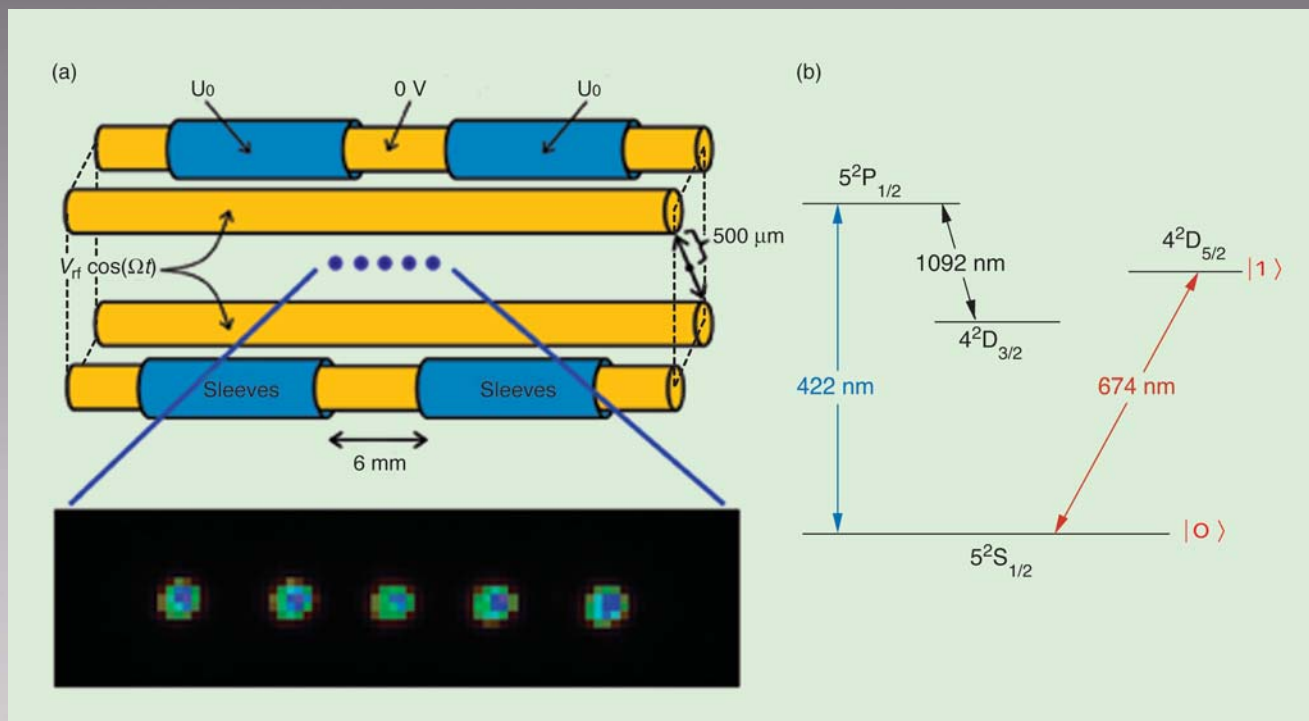


Figure 1. Strontium Ion Linear rf Paul Trap

(a) A schematic of the linear trap depicts five $^{88}\text{Sr}^+$ ions along its axis (not to scale). The ions in this trap are confined radially in a time-averaged potential that is created by applying 100 V at a frequency of 7 MHz to the two electrodes shown. The other two electrodes are held at a constant potential. The tubular electrodes (labeled “sleeves”) are held at constant potentials up to 100 V, relative to the other electrodes, to stop the ions from leaking out of the ends of the trap. The picture of five Sr^+ ions was made by focusing the 422-nm light scat-

tered from the ions onto an intensified charge-coupled-device camera. The ions are spaced about $20\ \mu\text{m}$ from each other. (b) The diagram shows the relevant energy levels of Sr^+ and the corresponding transitions (not to scale). We use 422-nm light from a frequency-doubled diode laser to Doppler-cool the ions and collect the scattered 422-nm light to detect the ions. A fiber laser generates 1092-nm light that keeps the ions from becoming stuck in the long-lived $\text{D}_{3/2}$ state. A very stable diode laser at 674 nm drives the narrow $\text{S}_{1/2} \leftrightarrow \text{D}_{5/2}$ transition.

so that quantum jumps can be observed in the experiments discussed next.

Quantum Randomness

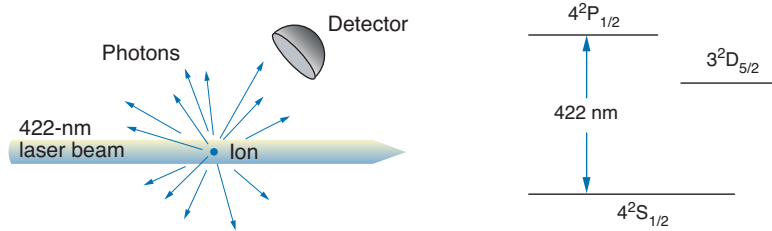
In the article “A New Face for Cryptography” on page 68, the

authors describe the quantum cryptography project at Los Alamos. Cryptography applications, whether classical or quantum, require strings of numbers (typically 1s and 0s) that are as random as possible. Generating random numbers, however, is not a trivial matter. In fact, the random number generators found in various

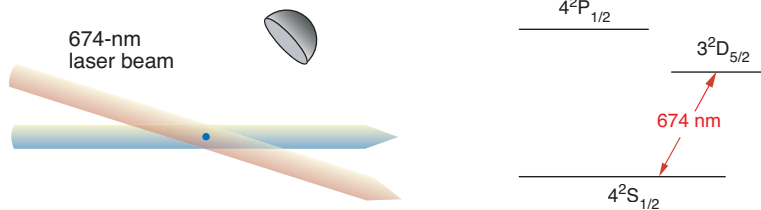
computer programs do not yield very random numbers because they are based on algebraic processes that are intrinsically deterministic.

It is generally accepted that producing strings of truly random numbers requires measuring the random outcome of a quantum-mechanical process. One example of a random

(a) An ion excited to the short-lived $4^2P_{1/2}$ state scatters millions of photons per second



(b) The scattering stops when the ion jumps to the long-lived $3^2D_{5/2}$ state



(c) Data from a quantum-jump experiment

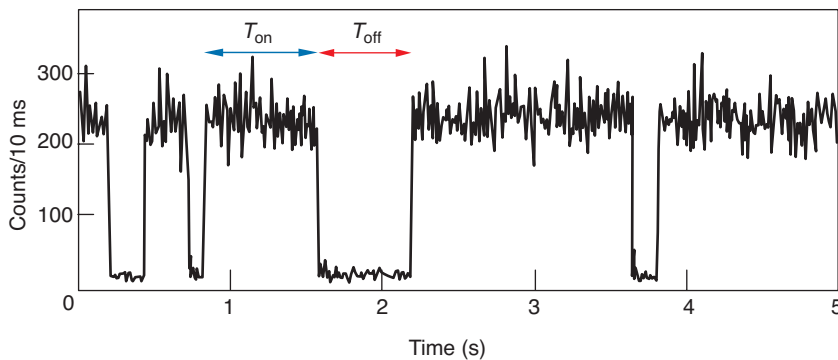


Figure 2. Quantum Jumps in a Single Trapped $^{88}\text{Sr}^+$ Ion

(a) When illuminated by 422-nm radiation, a single strontium ion will cycle between the $S_{1/2}$ and $P_{1/2}$ states and will scatter millions of photons per second. Some of the scattered light can be collected with a simple optical detector in order to monitor the state of the ion. (b) If the ion is simultaneously illuminated with 674-nm radiation, it will occasionally undergo a transition (“quantum jump”) from the $S_{1/2}$ state to the long-lived $D_{5/2}$ state. The scattered light then disappears. (c) This plot shows typical data from the quantum-jump experiment. When the count rate is over 50 counts per 10 ms, the atom is cycling between the $S_{1/2}$ and $P_{1/2}$ states. When the count rate suddenly falls to less than 50 counts per 10 ms, the atom has made a transition into the $D_{5/2}$ state. We continuously monitor the ion’s scattering rate for nearly an hour to observe tens of thousands of these transitions.

outcome is a photon hitting a beam splitter (Jennewein et al. 2000). The photon has a probability to either pass through the optic or reflect off it, and only a measurement determines its fate. Another example is the decay of radioactive nuclei, which emit, say,

alpha particles at unpredictable times (Silverman et al. 2000). Although both those processes are believed to be random, they suffer from one major drawback in a test of their statistics: As in any experimental setup, all the detectors have physical limitations.

Therefore, we cannot be sure that we would detect every photon or alpha particle. It is possible that some non-random processes might be overlooked in analyzing the incomplete data set.

In contrast, a very clean way to test the statistical nature of quantum processes is to analyze the behavior of an atom undergoing quantum jumps (Erber 1995). Quantum jumps are the sudden transitions from one quantum state to another. As Figure 2 shows, a strontium ion in the $S_{1/2}$ ground state will absorb a photon from a laser tuned to 422 nanometers and “jump” to the $P_{1/2}$ excited state. Because the $P_{1/2}$ state is short-lived, the ion quickly returns to the $S_{1/2}$ state by emitting a 422-nanometer photon in a random direction. Once it returns to the $S_{1/2}$ state, the ion can absorb and emit another photon, and because the lifetime of the $P_{1/2}$ excited state is so short, the ion will scatter millions of photons per second. We can detect enough of the scattered light with an optical system to observe the ion but not enough to determine every time the ion jumps to and from the $P_{1/2}$ state.

To directly observe quantum jumps, we simultaneously illuminate the ion with a 422- and a 674-nanometer laser light. In addition to jumping to the $P_{1/2}$ state, now the ion can also jump to the $D_{5/2}$ state. As soon as that transition occurs, the ion will stop scattering 422-nanometer light. The scattered light will return the moment the ion has left the $D_{5/2}$ state. As Figure 2 shows, we can very easily record every time a single ion makes a transition to the $D_{5/2}$ state and every time it returns to the $S_{1/2}$ state. According to quantum theory, the exact times of those transitions are completely unpredictable. Surprisingly, this prediction has not been tested with data sets comprising much more than about a thousand consecutive events. It is important to test very large sets of data because it

is harder to make a nonrandom series of numbers appear random if the series is very long.

Many tests can be used to determine the degree of randomness in a string of data. Figure 3(a) shows the result of one such test applied to our quantum-jump data (Itano et al. 1990). A single atom was continuously monitored until it had made over 34,000 transitions in and out of the $D_{5/2}$ state. We record the length of each time period $T_{\text{on},i}$, during which the atom continually scatters 422-nanometer photons, and the length of each subsequent time period $T_{\text{off},i}$, during which the ion scattered no photons because it was in the $D_{5/2}$ state. For example, in the figure, the values of T_{off} are $T_{\text{off},1} = 0.23$ second, $T_{\text{off},2} = 0.1$ second, $T_{\text{off},3} = 0.61$ second, and $T_{\text{off},4} = 0.17$ second.

We then sift through the data to determine the number of times a particular pair of values ($T_{\text{off},i}, T_{\text{off},i+1}$) occurs and make the color-coded plot shown in Figure 3(a). The symmetry and shapes of these graphs reflect several important characteristics of the data. For example, a pair of values, say $(T_{\text{off},i}, T_{\text{off},i+1}) = (0.23 \text{ second}, 0.1 \text{ second})$, is just as likely to occur as the pair $(0.1 \text{ second}, 0.23 \text{ second})$ —a long period of fluorescence is no more likely to be followed by a short one than a short period is likely to be followed by a long one. Essentially, plots like these indicate that the ion has no memory of what it was doing just the briefest moment before it fluoresces. This fundamental feature of quantum processes has not previously been tested precisely. It is also exactly what one would like to see in a random number generator.

We can easily convert the quantum-jump data into a string of 1s and 0s. If $T_{\text{on},i}$ is more than a set amount of time, we assign to that event the value 0. Likewise, if $T_{\text{on},i}$ is less than this time, we will assign the value 1

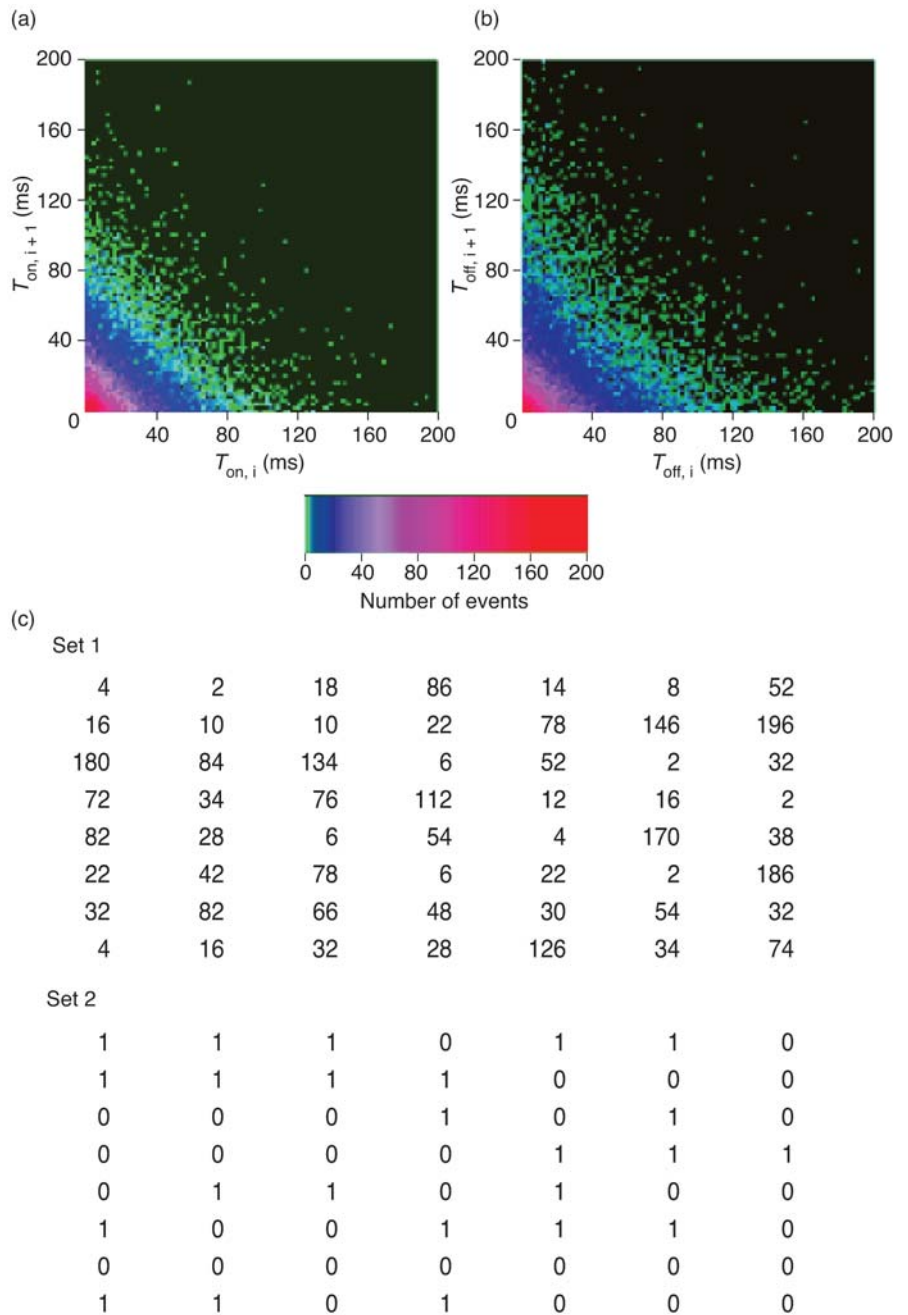


Figure 3. Analyzing Quantum-Jump Data

The scatter plots show consecutive periods that the ion spends (a) scattering 422-nm photons ($T_{\text{on},i}, T_{\text{on},i+1}$) and (b) not scattering 422-nm photons ($T_{\text{off},i}, T_{\text{off},i+1}$). Because these graphs are symmetric about their diagonal axis, we can tell that the ion is just as likely to spend a long time scattering photons followed by a short time scattering photons as it is to spend a short time followed by a long time scattering photons. This is one of many indications that the ion has no memory of when it has made a transition between the $S_{1/2}$ and $D_{5/2}$ states. (c) The quantum-jump data can also be converted to digital data. The first set of numbers shows a string of consecutive times spent in the $D_{5/2}$ state ($T_{\text{off},i}$). If the ion spends 30 ms or more in the $D_{5/2}$ state, the event is assigned a value of 0. Otherwise, the event is assigned a value of 1. These assignments are shown in set 2. With strings of tens of thousands of these digital numbers, we can use established protocols to test the randomness of our quantum-jump data.

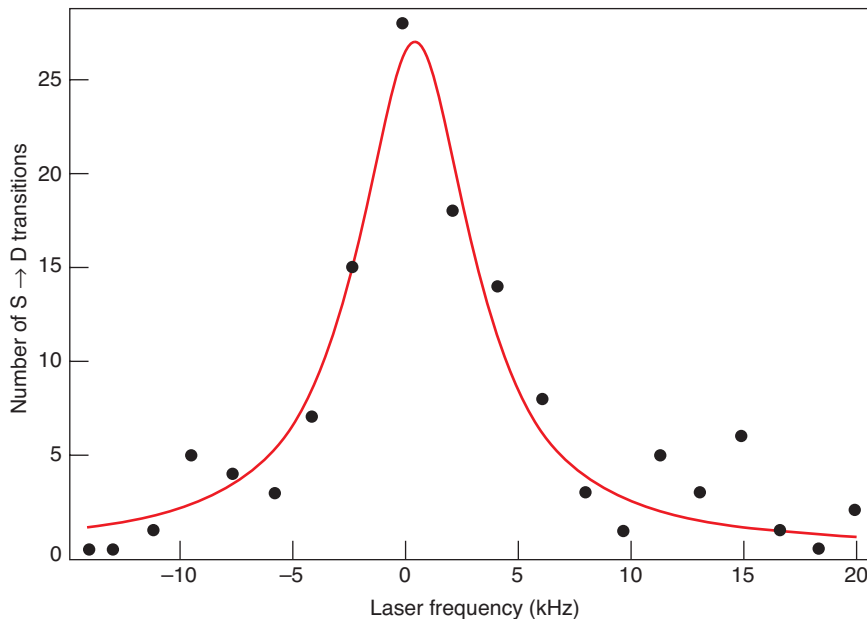


Figure 4. Measurement of the Laser Linewidth

The plot shows data taken from the narrow sideband of the $S_{1/2} \leftrightarrow D_{5/2}$ transition in a single trapped $^{88}\text{Sr}^+$ ion. The solid line is a Lorentzian line shape that is fitted to the data. In one laser probe cycle, the atom starts in the $S_{1/2}$ state. Next, the cooling light is turned off while the 674-nm light is pulsed on for 0.001s. Then the cooling light is turned on again, and we see if any 422-nm light is scattered into the detector. If not, then the 674-nm laser has successfully transferred the ion to the $D_{5/2}$ state. This process is repeated 100 times for each laser frequency.

to the event. Figure 3(b) gives an example of this conversion for a typical set of data.

Digitizing our data lets us use some of the established protocols that test the randomness of digital data. (One such standard is outlined in the U.S. Federal Information Processing Standards publication 140-2). An example of such a test is the following: In a string of 1s and 0s, we count how many times the two-digit patterns (0,0), (0,1), (1,0) and (1,1) appear. We then compare these numbers with the values expected for an ideal, random sequence. It is easy to calculate how likely it is that the measured sets of values differ from the expected ones, so that we can decide whether or not our quantum-jump data are random according to the given protocol. We are collecting continuous sequences of data, tens of thousands of events

long, that can be used for these tests.

Quantum Computing

We are also beginning some of the tasks that are prerequisites to making a quantum logic gate with a trapped ion. Perhaps the most critical step is coherently driving transitions between specific qubit states. In the experiments we are considering, the strontium $S_{1/2}$ ground state corresponds to the qubit state $|0\rangle$, whereas the $D_{5/2}$ excited state corresponds to the $|1\rangle$ qubit state. The stable 674-nanometer diode laser couples the qubit states to each other and to states of the ion's quantized external motion that would also be qubit states (Monroe et al. 1995).

The stability of the laser is one of several parameters that can limit the performance of a quantum com-

puter. If the laser frequency and phase were constant, we could almost always complete quantum logic operations perfectly. For example, starting with the ion in the $S_{1/2}$ state, we could reliably create a specific superposition of the $S_{1/2}$ and $D_{5/2}$ states:

$$|S_{1/2}\rangle = \frac{|S_{1/2}\rangle + i|D_{5/2}\rangle}{\sqrt{2}}. \quad (1)$$

However, if the phase or frequency of the laser is not perfectly stable while this operation is taking place, the result of the operation may be, for example,

$$|S_{1/2}\rangle = \frac{0.99|S_{1/2}\rangle + i1.01|D_{5/2}\rangle}{\sqrt{2}} \quad (2)$$

In this case, the new wave function has a small phase error. If this operation is repeated many times, the accumulations of these small errors could invalidate the results of a quantum computation. Because every laser has a nonzero linewidth (proportional to the laser's frequency), such errors are inevitable. One way to reduce the likelihood of introducing the errors is to perform the logic operation quickly, that is, faster than the typical time scales of the frequency fluctuations of the laser, although it is easier to perform a quantum-gate operation slowly. Thus, it is critical that the laser be very stable with its linewidth as small as possible.

We have measured our laser linewidth using a procedure related to the quantum-jump experiment described earlier. First, we turn off the 422-nanometer light, letting the ion decay to the $S_{1/2}$ state. Then we illuminate the ion with a pulse of 674-nanometer laser light. (The 422-nanometer light remains off during this step, because that light will perturb the $S_{1/2}$ state and broaden the $S_{1/2} \leftrightarrow D_{5/2}$ transition.) We then

determine whether or not the laser has driven the atom from the $S_{1/2}$ to the $D_{5/2}$ state by shining the 422-nanometer light on the ion. We detect light scattered by the ion if it is not in the $D_{5/2}$ state, but only background light (the small amount of light scattered off the trap and vacuum chamber) if the ion is in the $D_{5/2}$ state. Figure 4 shows the number of times the 674-nanometer laser transfers the ion to the $D_{5/2}$ state as the laser frequency is scanned over one of the motional sidebands of the $S_{1/2} \leftrightarrow D_{5/2}$ transition. The figure also shows the result of fitting a Lorentzian-shaped curve to these data. From the shape of the fitted curve and from a few key experimental parameters, we can determine that the laser linewidth is about 4 kilohertz or less, which is about one percent of one billionth of the absolute frequency of the laser light (445 terahertz).

This laser linewidth is sufficiently narrow so that we can perform specific, coherent operations on qubit states. However, to perform the operations needed for a quantum logic gate, the ions must be cooled much more than they are at present, so that the quantum state of the ion can be initialized to the ground state of its motion. We are currently working toward this goal and on further narrowing the linewidth of the 674-nanometer laser. In addition, we are working on or anticipate performing several other quantum-optics experiments. The apparatus presented here, along with ion traps in general, can facilitate significant contributions to the field of quantum information and quantum computation. ■

Acknowledgments

I would like to thank Richard Hughes for suggesting the study of the randomness of quantum jumps and for his indispensable role in bringing ion-trap technology to the field of quantum information and quantum computation at Los Alamos. In addition, I am grateful to Daisy Raymondson for her work on some of the laser systems used in this experiment and to the Los Alamos Summer School for initially bringing her to Los Alamos.

Further Reading

- Berkeland, D. J., A Linear Paul Trap for Strontium Ions. (To be published in *Rev. Sci. Instrum.*)
- Berkeland, D. J., J. D. Miller, F. C. Cruz, B. C. Young, R. J. Rafac, X.-P. Huang et al. 1999. High-Resolution, High-Accuracy Spectroscopy of Trapped Ions. In *Atomic Physics 16. Proceedings of the International Conference*. Edited by W. E. Baylis and G. W. F. Drake, 29. New York: AIP Press.
- Erber, T. 1995. Testing the Randomness of Quantum-Mechanic: Natures Ultimate Cryptogram. *Ann. N. Y. Acad. Sci.* **755**: 748.
- Itano, W. M., J. C. Bergquist, F. Diedrich, and D. J. Wineland. 1990. Quantum Optics of Single, Trapped Ions. In *Coherence and Quantum Optics VI*, 539. Edited by J. H. Eberly, et al. New York: Plenum Press.
- Jenneweine, T., U. Achleitner, G. Weihs, H. Weinfurter, and A. Zeilinger. 2000. A Fast and Compact Quantum Random Number Generator. *Rev. Sci. Instrum.* **71** (4): 1675.
- Monroe, C., D. M. Meekhof, B. E. King, W. M. Itano, and D. J. Wineland. 1995. Demonstration of a Fundamental Quantum Logic Gate. *Phys. Rev. Lett.* **75**: 4714.
- Silverman, M. P., W. Strange, C. Silverman, and T. C. Lipscombe. 2000. Tests for Randomness of Spontaneous Quantum Decay. *Phys. Rev. A* **61**: 042106.
- Thompson, R. C., K. Dholakia, J.-L. Hernandez-Pozos, G. Zs. K. Horvath, J. Rink, and D. M. Segal. 1997. Spectroscopy and Quantum Optics with Ion Traps. *Physica Scr.* **T72**: 24.

Dana J. Berkeland has performed a wide variety of experiments in atomic and optical physics. She received her Ph.D. from Yale University in 1995 after precisely measuring Stark shifts in

lithium and cesium and measuring the ground-state Lamb shift in hydrogen to an accuracy of 6 parts per million. This measurement, performed with Malcolm Boshier, was at the time the most precise measurement of the quantity.

Dana spent three years as a National Research Council postdoctoral fellow at the National Institute of Standards and Technology in Boulder, Colorado, where she evaluated a trapped-mercury-ion microwave frequency standard to a fractional accuracy of 3×10^{-15} (making it one of the most accurate frequency standards in the world). In 1998, she came to Los Alamos National Laboratory as a J. Robert Oppenheimer postdoctoral fellow to build a laboratory for quantum-optics experiments with trapped strontium ions and became a technical staff member in 2000. She has over a decade of experience in building and developing laser systems and related optics. Dana has over seven years of experience in building and working with ion trapping systems.



Theory of Single-Spin Detection with a Scanning Tunneling Microscope

Alexander V. Balatsky and Ivar Martin

No fundamental principle precludes the measurement of a single spin, and therefore the capability to make such a measurement simply depends on our ability to develop a detection method of sufficient spatial and temporal resolution. The standard electron spin detection technique—electron spin resonance—is limited to a macroscopic number of electron spins (10^{10} or more) (Farle 1998). A state-of-the-art magnetic resonance force microscope has recently detected about a hundred fully polarized electron spins (Bruland et al. 1998). We argue that scanning tunneling microscopy offers a powerful technique to detect a single spin and propose the theoretical basis for the new spin-detection technique, which we call spin precession by scanning tunneling microscopy.

The capability to routinely detect and manipulate a single spin would be remarkably useful, with applications ranging from the study of strongly correlated systems to nanotechnology and quantum information processing. For example, we could investigate magnetism on the nanoscale in a strongly correlated system by detecting changes in the spin behavior as the system enters the magnetically ordered state (Heinze et al. 2000). We could also fully explore the magnetic properties of a single paramagnetic atom in the Kondo regime (Manoharan et al. 2000). Magnetic properties of spin centers in superconductors are another area where a single spin plays an important role, since it can generate intragap impurity states (Salkola et al. 1997, Yazdani et al. 1997). With regard to nanotechnology, the ability to manipulate a single spin could open the door to single-spin-based information storage devices, whereas in the realm of quantum computing, it could help bring to fruition several specific computing architectures (Kane 1998, Loss and DiVincenzo 1998).

Our theoretical investigation of spin precession—scanning tunneling microscopy has in part been motivated by the experiments of Yshay Manassen et al. (1989), in which a defect structure (an oxygen vacancy) in oxidized silicon was interrogated with a scanning tunneling microscope (STM). The STM operated in the presence of an external magnetic field, and a small alternating current (ac) signal in the power spectrum of the tunneling current was detected at the spin's precession, or Larmor, frequency. The ac signal was spatially localized at distances of about 5–10 angstroms from the spin site. The extreme localization of the signal and the linear scaling of its frequency with the magnetic field prompted Manassen to attribute the detected ac signal to the Larmor precession of a single-spin site. Whereas that interpretation was somewhat controversial, the later work by Manassen et al. (2000) and more recent work by Colm Durkan and Mark Welland (2002) support the notion that STM can indeed sense a single spin.

From a theoretical perspective, it was not clear how the spin could generate an ac component in the STM's tunneling current. As outlined below, however, the precessing spin causes an ac modulation of the surface density of states near the spin site, provided a dc current flows through the surface. In fact, that current can be the tunneling current that flows between the STM tip and the surface. Thus, the tunneling current, which is proportional to the surface density of states, plays two roles in spin detection by scanning tunneling microscopy: It provides a means to couple the precessing spin to

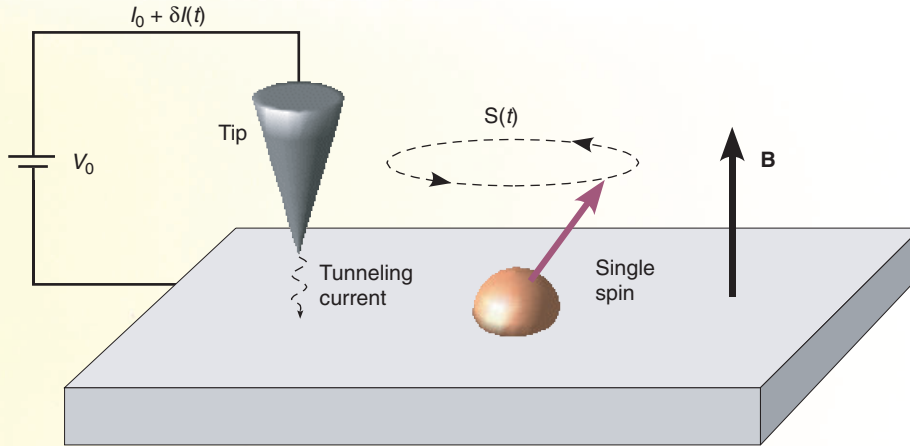


Figure 1. Experimental Setup for Electron Spin Precession by Scanning Tunneling Microscopy
In the applied magnetic field B , the spin of the magnetic atom (for example, gadolinium, shown in gold) is precessing around the field line. The STM tip is precisely positioned within a few angstroms of the spin site. The dc tunneling current I_0 , between the STM tip and the sample, can acquire an ac component, $\delta I(t)$, that signals the presence of the precessing spin.

the density of states and a means to detect the ac modulation of those states. The experimental setup that we consider is shown in Figure 1. A general discussion of the principles underlying scanning tunneling microscopy can be found on page 303.

Before analyzing the effect of the STM, consider a localized magnetic site with spin S (spin $1/2$), on the surface of a substrate. In the presence of a magnetic field, B , the energy levels of the spin-up and spin-down states (denoted by E_{\uparrow} and E_{\downarrow} , respectively) are Zeeman-split. At a finite temperature, or as a result of an external excitation, the spin may be driven into the mixed state characterized by the wave function

$$|\psi(t)\rangle = \alpha(t) |\uparrow\rangle + \beta(t) |\downarrow\rangle, \quad (1)$$

where

$$\alpha(t) = |\alpha| \exp(-iE_{\uparrow}t), \text{ and} \\ \beta(t) = |\beta| \exp(-iE_{\downarrow}t + i\phi(t)).$$

The phase $\phi(t)$ determines the spin coherence time τ_{ϕ} and is related to the spin relaxation time T_2 measured by electron spin resonance.

In the state given by Equation (1), the spin, with an expectation value of

$$\frac{\langle \psi(t) | \mathbf{S} | \psi(t) \rangle}{\langle \psi(t) | \psi(t) \rangle}, \quad (2)$$

will precess around a magnetic field line at the Larmor frequency ω_L ,

$$\hbar \omega_L = E_{\uparrow} - E_{\downarrow} = \gamma B \quad (3)$$

where γ is the gyromagnetic ratio. (See the box ‘‘Spin Manipulation with Magnetic Resonance’’ on page 288.) In a magnetic field of 100 gauss, this frequency is 280 megahertz for a free electron.

If we consider what happens on the surface, then the precession of the local moment will be coupled to the orbital motion of electrons via the spin-orbit interaction. The details of the spin-orbit coupling depend on the specific material. In general, however,

the interaction of the conduction electrons with the local impurity spin can be described by the Hamiltonian

$$H = H_0 + J \mathbf{S} \cdot \boldsymbol{\sigma}(0) , \quad (4)$$

where J is the strength of the exchange interaction between the local spin \mathbf{S} , and the spin density of the conduction electrons, $\boldsymbol{\sigma}(0) = \sigma_{\alpha\beta} c_{\alpha}^{\dagger}(0) c_{\beta}(0)$, on the impurity site. Here, $c_{\alpha}^{\dagger}(0)$, $c_{\beta}(0)$ are the electron creation/destruction operators with spin α and β , respectively, and $\sigma_{\alpha\beta} = (\sigma_{\alpha\beta}^x, \sigma_{\alpha\beta}^y, \sigma_{\alpha\beta}^z)$ is a vector of Pauli matrices. The unperturbed Hamiltonian H_0 describes the surface without the spin impurity. Based on symmetry, the energy of the unperturbed surface states contains a spin-orbit part that is linear both in the conduction-electrons' spin, $\boldsymbol{\sigma}$, and their momentum, \mathbf{k} (Bychkov and Rashba 1984).

$$\varepsilon(\mathbf{k}) = \frac{k^2}{2m^*} + \gamma_{\text{SO}} [\mathbf{n} \times \mathbf{k}] \cdot \hat{\boldsymbol{\sigma}} , \quad (5)$$

where m^* is the band mass of electrons in the substrate, \mathbf{n} is a unit vector normal to the surface, and γ_{SO} is a parameter that characterizes the strength of the surface spin-orbit coupling. The problem specified by Equations (4) and (5) can be solved for each instantaneous value of the precessing spin $\mathbf{S}(t)$. The solution, however, does not lead to a time-dependent conduction-electron density of states $N(\mathbf{r}, t)$ because the effects of the precessing spin average to zero. In that case, the tunneling current would remain constant.

To extend the model, we account for the fact that the tunneling current injects electrons into the sample, and those electrons can flow to the spin site. In the presence of a current density \mathbf{j} flowing through the surface, the equilibrium momentum distribution \mathbf{k} is shifted by an amount, $\mathbf{k}_0 = \mathbf{j}m^*/ne$, where n is the carrier density and e is the electron charge. This shift can be introduced into a Green's function matrix for the conduction electrons, $\hat{G}_0(\mathbf{k}, \omega)$,

$$\hat{G}_0(\mathbf{k}, \omega) = \left[\omega - \frac{(\mathbf{k} - \mathbf{k}_0)^2}{2m^*} - \gamma_{\text{SO}} [\mathbf{n} \times \mathbf{k}] \cdot \hat{\boldsymbol{\sigma}} \right]^{-1} . \quad (6)$$

We expand the matrix in γ_{SO} relative to the Fermi energy. Then, to first order in both the exchange coupling J and γ_{SO} , we obtain an \mathbf{S} -dependent contribution to the density of the surface states:

$$\frac{\delta N}{N} = \gamma_{\text{SO}} J \frac{dN}{dE} J_0^2(k_F r) [\mathbf{k}_0 \times \mathbf{S}]_n . \quad (7)$$

This correction depends on the distance from the spin center, r , through the Bessel function of the first kind, $J_0(x)$. The correction is time dependent in the presence of a magnetic field because the projection of \mathbf{S} oscillates at the Larmor frequency. The magnitude of the correction is proportional to the current density in the system (through \mathbf{k}_0).

The total (ac plus dc) tunneling current I , between the STM tip and the sample is proportional to the single-electron density of states in the substrate. Therefore, the

ac component $\delta I(t)$, normalized to the tunneling current, can be estimated as

$$\frac{\delta I(t)}{I} = \frac{\delta N(t)}{N}. \quad (8)$$

We have focused on the case in which an STM injects current into the system, but in principle, the current can also be provided externally (through extra leads attached to the substrate), and the ac current can be detected with some ultrasensitive current measurement device.

It is also important to note that the electron density of states $N(\mathbf{r}, t)$ is a scalar and should be invariant under time reversal, whereas \mathbf{S} is odd under time reversal. Hence, $\delta N(\mathbf{r}, t)$ can depend only on the product of the spin vector with some other vector that is odd under time reversal. In Equation (7), that vector is the current density, that is, $\delta N \sim [\mathbf{k}_0 \times \mathbf{S}]_n$. Another possibility is that the correction to the density of states depends on the time derivative of the spin vector, that is, $\delta N \sim \partial_t \mathbf{S}(t)$. We have also found a mechanism for this possibility.

Our conjecture of how an STM can detect single spins is based on the ac modulation of the density of surface states that results from a current-induced spin-orbit coupling to the precessing local spin. The changing state density is observed as the ac component to the tunneling current. ■

Further Reading

- Balatsky, A. V., and I. Martin. 2001. Theory of Single Spin Detection with STM. [Online]: <http://eprints.lanl.gov> (cond-mat/0112407).
- Bruland, K. J., W. M. Dougherty, J. L. Garbini, J. A. Sidles, and S. H. Chao. 1998. Force-Detected Magnetic Resonance in a Field Gradient of 250 000 Tesla per Meter. *Appl. Phys. Lett.* **73** (21): 3159.
- Bychkov, Y. A., and E. I. Rashba. 1984. Properties of a 2D Electron Gas with Lifted Spectral Degeneracy. *JETP Lett.* **39** (2): 78.
- Durkan, C., and M. E. Welland. 2002. Electronic Spin Detection in Molecules Using Scanning-Tunneling-Microscopy-Assisted Electron-Spin Resonance. *Appl. Phys. Lett.* **80** (3): 458.
- Farle, M. 1998. Ferromagnetic Resonance of Ultrathin Metallic Layers. *Rep. Prog. Phys.* **61** (7): 755.
- Heinze, S., M. Bode, A. Kubetzka, O. Pietzsch, X. Nie, S. Blugel, and R. Wiesendanger. 2000. Real-Space Imaging of Two-Dimensional Antiferromagnetism on the Atomic Scale. *Science* **288** (5472): 1805.
- Kane, B. E. 1998. A Silicon-Based Nuclear Spin Quantum Computer. *Nature* **393**: 133.
- Loss, D., and D. P. DiVincenzo. 1998. Quantum Computation with Quantum Dots. *Phys. Rev. A* **57**: 120.
- Manassen, Y., I. Mukhopadhyay, and N. R. Rao. 2000. Electron-Spin-Resonance STM on Iron Atoms in Silicon. *Phys. Rev. B* **61** (23): 16223.
- Manassen, Y., R. J. Hamers, J. E. Demuth, and A. J. Castellano Jr. 1989. Direct Observation of the Precession of Individual Paramagnetic Spins on Oxidized Silicon Surfaces. *Phys. Rev. Lett.* **62**: 2531.
- Manoharan, H. C., C. P. Lutz, and D. M. Eigler. 2000. Quantum Mirages Formed by Coherent Projection of Electronic Structure. *Nature* **403**: 512.
- Salkola, M. I., A. V. Balatsky, and J. R. Schrieffer. 1997. Spectral Properties of Quasiparticle Excitations Induced by Magnetic Moments in Superconductors. *Phys. Rev. B* **55**: 12648.
- Wiesendanger, R., H.-J. Güntherodt, G. Güntherodt, R. J. Gambino, and R. Ruf. 1990. Observation of Vacuum Tunneling of Spin-Polarized Electrons with the Scanning Tunneling Microscope. *Phys. Rev. Lett.* **65**: 247.
- Yazdani, A., B. A. Jones, C. P. Lutz, M. F. Crommie, and D. M. Eigler. 1997. Probing the Local Effects of Magnetic Impurities on Superconductivity. *Science* **275** (5307): 1767.

Alexander Balatsky received his Ph.D. in 1987 from the Landau Institute for Theoretical Physics,

where he then worked as a researcher until 1989. From 1989 to 1991, he was at the University of Illinois at



Urbana-Champaign, where he became a visiting resident assistant professor in 1990. In 1991, he joined Los Alamos National Laboratory as a J. R. Oppenheimer Fellow and is currently a technical staff member in the Theoretical Division.

Ivar Martin is a technical staff member in the Theoretical Division at Los Alamos National Laboratory.

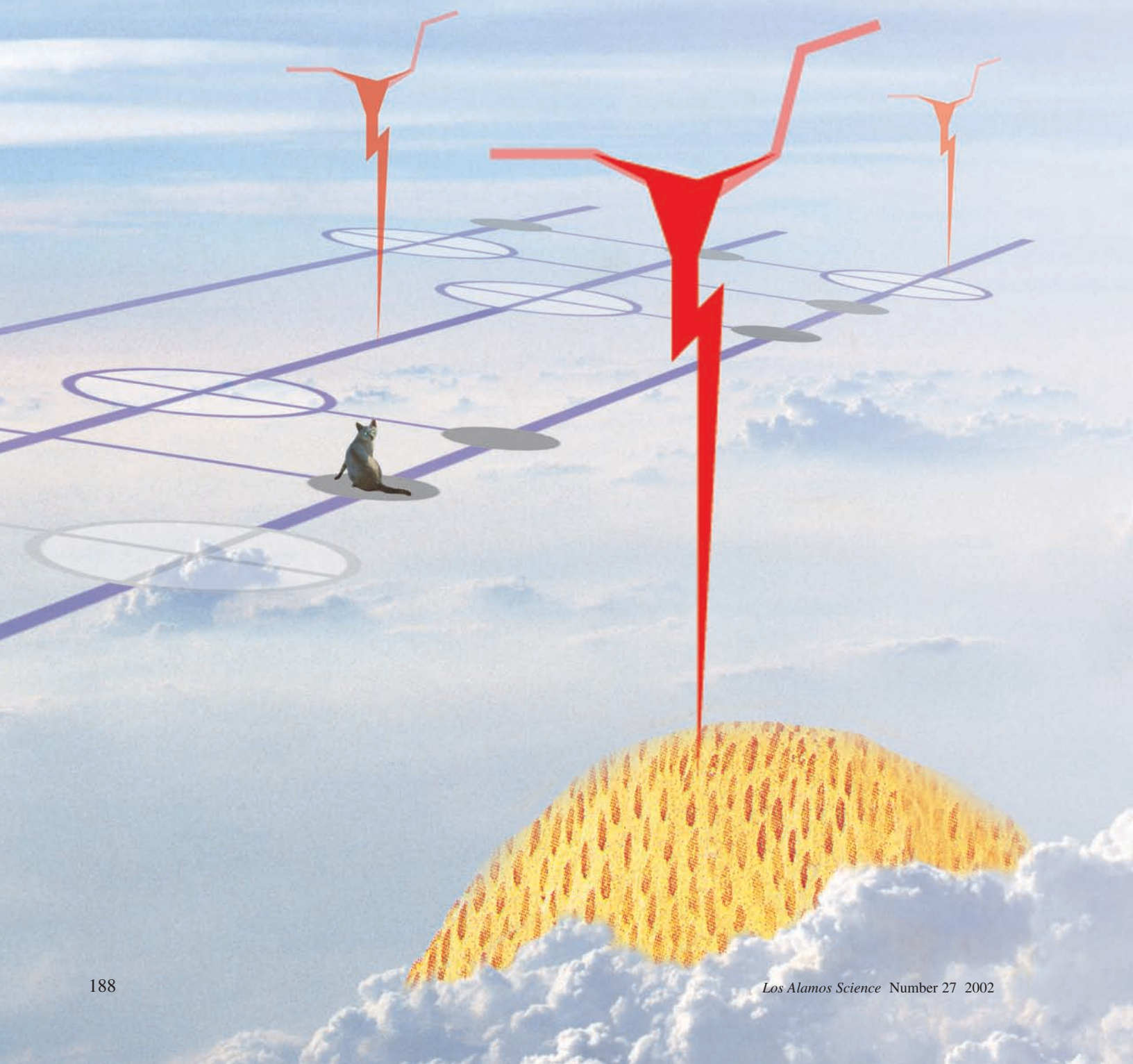
He received his Ph.D. from the University of Illinois at Urbana-Champaign in 1999. His research interests include



the theory of strongly correlated systems, development of novel local probes, and the theory of quantum measurement and computation.

Introduction to Quantum Error Correction

*Emanuel Knill, Raymond Laflamme, Alexei Ashikhmin, Howard N. Barnum,
Lorenza Viola, and Wojciech H. Zurek*



When physically realized, quantum information processing (QIP) can be used to solve problems in physics simulation, cryptanalysis, and secure communication for which there are no known efficient solutions based on classical information processing. Numerous proposals exist for building the devices required for QIP by using systems that exhibit quantum properties. Examples include nuclear spins in molecules, electron spins or charge in quantum dots, collective states of superconductors, and photons (Braunstein and Lo 2000). In all these cases, there are well-established physical models that, under ideal conditions, allow for exact realizations of quantum information and its manipulation. However, real physical systems never behave exactly like the ideal models. The main problems are environmental noise, which is due to incomplete isolation of the system from the rest of the world, and control errors, which are caused by calibration errors and random fluctuations in control parameters. Attempts to reduce the effects of these errors are confronted by the conflicting needs of being able to control and reliably measure the quantum systems. These needs require strong interactions with control devices and systems that are sufficiently well isolated to maintain coherence, the subtle relationship between the phases in a quantum superposition. The fact that quantum effects rarely persist on macroscopic scales suggests that meeting these needs requires considerable outside intervention.

Soon after Peter Shor published the efficient quantum factoring algorithm with its applications to breaking commonly used public-key cryptosystems, Andrew Steane (1996) and Shor (1995) gave the first constructions of quantum error-correcting codes. These codes make it possible to store quantum information so that one can reverse the effects of the most likely errors. By demonstrating that quantum information can exist in protected parts of the state space, they showed that, in principle, it is possible to protect against environmental noise when storing or transmitting information. Stimulated by these results and in order to solve errors happening during computation with quantum information, researchers initiated a series of investigations to determine whether it was possible to quantum-compute in a fault-tolerant manner. The outcome of these investigations was positive and culminated in what are now known as accuracy threshold theorems (Gottesman 1996, Calderbank et al. 1997, Calderbank et al. 1998, Shor 1996, Kitaev 1997, Knill and Laflamme 1996, Aharonov and Ben-Or 1996, Aharonov and Ben-Or 1999, Knill et al. 1998a, Knill et al. 1998b, Gottesman 1998, Preskill 1998). According to these theorems, if the effects of all errors are sufficiently small per quantum bit (qubit) and computation step, then it is possible to process quantum information arbitrarily accurately with reasonable resource overheads. The requirement on errors is quantified by a maximum tolerable error rate called the threshold. The threshold value depends strongly on the details of the assumed error model. All threshold theorems require that errors at different times and locations be independent and that the basic computational operations can be applied in parallel. Although the proven thresholds are well out of the range of today's devices, there are signs that, in practice, fault-tolerant quantum computation may be realizable.

In retrospect, advances in quantum error correction and fault-tolerant computation were made possible by the realization that accurate computation does not require the state of the physical devices supporting the computation to be perfect. In classical information processing, this observation is so obvious that it is often forgotten: No two letters "e" on a written page are physically identical, and the number of electrons used to store a bit in the computer's memory varies substantially. Nevertheless, we have no difficulty in accurately identifying the desired letter or state. A crucial conceptual difficulty with quantum information is that, by its very nature, it cannot be identified by being "looked" at. As a result, the sense in which quantum information can be

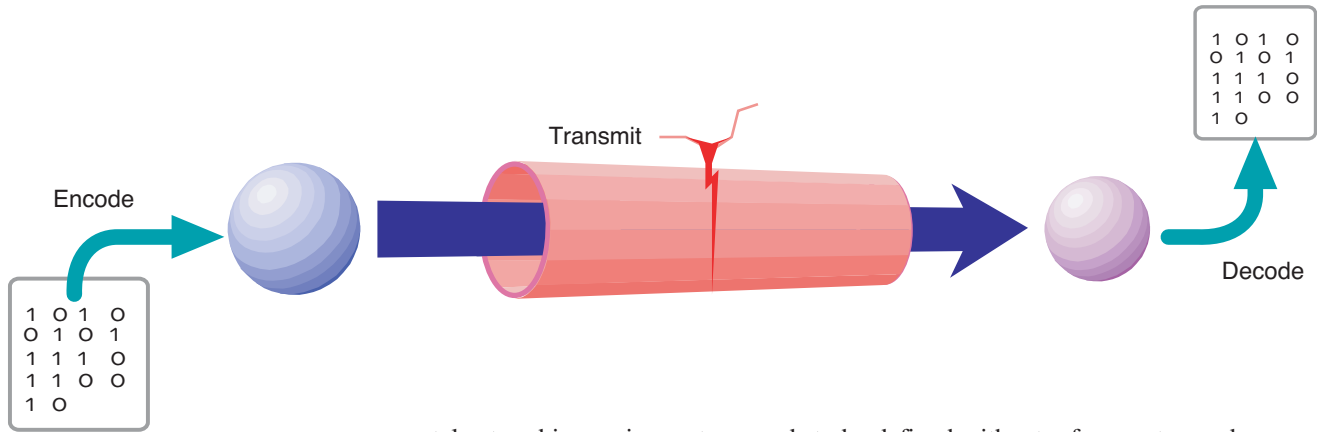


Figure 1. Typical Application of Error Correction Methods

The three main steps required for communication are shown in this figure: Information is first encoded in a physical system, then transmitted over the noisy communication channel, and finally decoded. The combination of encoding and decoding is chosen so that errors have no effect on the transmitted information.

accurately stored in a noisy system needs to be defined without reference to an observer. There are two ways to accomplish this task. The first is to define stored information to be the information that can, in principle, be extracted by a quantum decoding procedure. The second is to explicitly define “subsystems” (particle-like aspects of the quantum device) that contain the desired information. The first approach is a natural generalization of the usual interpretations of classical error-correction methods, whereas the second is motivated by a way of characterizing quantum particles.

In this article, we motivate and explain the decoding and subsystems view of quantum error correction. We explain how quantum noise in QIP can be described and classified and summarize the requirements that need to be satisfied for fault tolerance. Considering the capabilities of currently available quantum technology, the requirements appear daunting. But the idea of subsystems shows that these requirements can be met in many different, and often unexpected, ways.

Our article is structured as follows: The basic concepts are introduced by example, first for classical and then for quantum codes. We then show how the concepts are defined in general. Following a discussion of error models and analysis, we state and explain the necessary and sufficient conditions for detectability of errors and correctability of error sets. That section is followed by a brief introduction to two of the most important methods for constructing error-correcting codes and subsystems. For a basic overview, it suffices to read the beginnings of these more-technical sections. The principles of fault-tolerant quantum computation are outlined in the last section.

Concepts and Examples

Communication is the prototypical application of error correction methods. To communicate, a sender needs to convey information to a receiver over a noisy communication channel. Such a channel can be thought of as a means of transmitting an information-carrying physical system from one place to another. During transmission, the physical system is subject to disturbances that can affect the information carried. To use a communication channel, the sender needs to encode the information to be transmitted in the physical system. After transmission, the receiver decodes the information. The procedure is shown in Figure 1.

Protecting stored information is another important application of error correction methods. In this case, the user encodes the information in a storage system and retrieves it later. Provided that there is no communication from the receiver to the sender, any error correction method applicable to communication is also applicable to storage and vice versa. In a later section (“Fault-Tolerant Quantum Communication and Computation” on page 217), we discuss the problem of fault-tolerant computation,

which requires enhancing error correction methods in order to enable applying operations to encoded information without losing protection against errors.

To illustrate the different features of error correction methods, we consider three examples. We begin by describing them for classical information, but in each case, there is a quantum analogue that will be introduced later.

Trivial Two-Bit Example. Consider a physical system consisting of two bits with state space $\{00, 01, 10, 11\}$. We use the convention that state symbols for physical systems subject to errors are in gray. States changed by errors are shown in red.¹ In this example, the system is subject to errors that flip (apply the **not** operator to) the first bit with probability .5. We wish to safely store one bit of information. To this end, we store the information in the second physical bit because this bit is unaffected by the errors (see Figure 2).

As suggested by the usage examples in Figure 1, one can encode one bit of information in the physical system by the map that takes $0 \rightarrow 00$ and $1 \rightarrow 01$. This means that the states 0 and 1 of an ideal bit are represented by the states 00 and 01 of the noisy physical system, respectively.

To decode the information, one can extract the second bit by the following map:

$00 \rightarrow 0$
 $10 \rightarrow 0$
 $01 \rightarrow 1$
 $11 \rightarrow 1$

This procedure ensures that the encoded bit is recovered by the decoding regardless of the error. There are other combinations of encoding and decoding that work. For example, in the encoding, we could swap the meaning of 0 and 1 by using the map $0 \rightarrow 01$ and $1 \rightarrow 00$. The new decoding procedure adds a bit flip to the one shown above. The only difference between this combination of encoding/decoding and the previous one lies in the way in which the information is represented in the range of the encoding. This range consists of the two states 00 and 01 and is called the code. The states in the code are called code words.

Although trivial, the example just given is typical of ways for dealing with errors. That is, there is always a way of viewing the physical system as a pair of abstract systems: The first member of the pair experiences the errors, and the second carries the information to be protected. The two abstract systems are called subsystems of the physical system and are usually not identifiable with any of the system's physical components. The first is the syndrome subsystem, and the second is the information-carrying subsystem. Encoding consists of initializing the first system and storing the information in the second. Decoding is accomplished by extraction of the second system. In the example, the two subsystems are readily identified as the two physical bits that make up the physical system. The first is the syndrome subsystem and is initialized to 0 by the encoding. The second carries the encoded information.

¹ These graphical conventions are not crucial for understanding what the symbols mean and are intended for emphasis only.

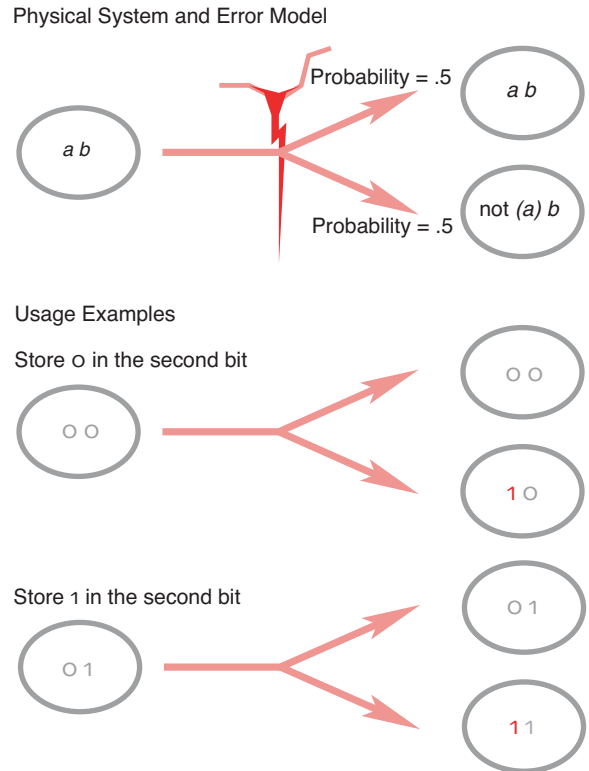


Figure 2. A Simple Error Model
 Errors affect only the first bit of a physical two-bit system. All joint states of the two bits are affected by errors. For example, the joint state 00 is changed by the error to 10 . Nevertheless, the value of the information represented in the second physical bit is unchanged.

The Repetition Code. The next example is a special case of the main problem of classical error correction and occurs in typical communication settings and in computer memories. Let the physical system consist of three bits. The effect of the errors is to independently flip each bit with probability p , which we take to be $p = .25$. The repetition code results from triplicating the information to be protected. An encoding is given by the map $0 \rightarrow 000, 1 \rightarrow 111$. The repetition code is the set $\{000, 111\}$, which is the range of the encoding. The information can be decoded with majority logic: If two out of three bits are 0, the output is 0; otherwise, the output is 1.

How well does this encoding/decoding combination work for protecting one bit of information against the errors? The decoding fails to extract the bit of information correctly if two or three of the bits were flipped by the error. We can calculate the probability of incorrect decoding as follows: The probability of a given pair of bits having flipped is $.25^2 * .75$. There are three different pairs. The probability of three bits having flipped is $.25^3$. Thus, the probability of error in the encoded bit is $3 * .25^2 * .75 + .25^3 = 0.15625$. This is an improvement over .25, which is the probability that the information represented in one of the three physical bits is corrupted by error.

To see that one can interpret this example by viewing the physical system as a pair of subsystems, it suffices to identify the physical system's states with the states of a suitable pair. The following shows such a subsystem identification:

$$\begin{array}{lcl}
 000 & \leftrightarrow & 00 \cdot 0 \\
 001 & \leftrightarrow & 11 \cdot 0 \\
 010 & \leftrightarrow & 01 \cdot 0 \\
 100 & \leftrightarrow & 10 \cdot 0 \\
 011 & \leftrightarrow & 10 \cdot 1 \\
 101 & \leftrightarrow & 01 \cdot 1 \\
 110 & \leftrightarrow & 11 \cdot 1 \\
 111 & \leftrightarrow & 00 \cdot 1
 \end{array} \tag{2}$$

The left side consists of the 8 states of the physical system, which are the possible states for the three physical bits making up the system. The right side shows the corresponding states for the subsystem pair. The syndrome subsystem is a two-bit subsystem, whose states are shown first. The syndrome subsystem's states are called syndromes. After the “.” symbol are the states of the information-carrying one-bit subsystem.

In the subsystem identification above, the repetition code consists of the two states for which the syndrome is 00. That is, the code states 000 and 111 correspond to the states 00 · 0 and 00 · 1 of the subsystem pair. For a state in this code, single-bit flips do not change the information-carrying bit, only the syndrome. For example, a bit flip of the second bit changes 000 to 010, which is identified with 01 · 0. The syndrome has changed from 00 to 01. Similarly, this error changes 111 to 101 \leftrightarrow 01 · 1. The following diagram shows these effects :

$$\begin{array}{lclclcl}
 000 & \leftrightarrow & 00 \cdot 0 & & 111 & \leftrightarrow & 00 \cdot 1 \\
 \downarrow & & & & \downarrow & & \\
 010 & \leftrightarrow & 01 \cdot 0 & & 101 & \leftrightarrow & 01 \cdot 1
 \end{array} \tag{3}$$

Note that the syndrome change is the same. In general, with this subsystem identification, we can infer from the syndrome which single bit was flipped on an encoded state.

Errors usually act cumulatively over time. For the repetition code, this is a problem in the sense that it takes only a few actions of the above error model for the two- and three-bit errors to overwhelm the encoded information. One way to delay the loss of information is to decode and reencode sufficiently often. Instead of explicitly decoding and reencoding, the subsystem identification can be used directly for the same effect, namely, that of resetting the syndrome subsystem's state to 00 . For example, if the state is $10 \cdot 1$, it needs to be reset to $00 \cdot 1$. Therefore, using the subsystem identification, resetting requires changing the state 011 to 111 . It can be checked that, in every case, what is required is to set all bits of the physical system to the majority of the bits. After the syndrome subsystem has been reset, the information is again protected against the next one-bit error.

A Code for a Cyclic System. We next consider a physical system that does not consist of bits. This system has seven states symbolized by $0, 1, 2, 3, 4, 5,$ and 6 . Let s_1 be the right-circular shift operator defined by $s_1(l) = l + 1$ for $0 \leq l \leq 5$ and $s_1(6) = 0$. Define $s_0 = \mathbb{1}$ (the identity operator),

$$s_k = \underbrace{s_1 \dots s_1}_{k \text{ times}},$$

and $s_{-k} = s_k^{-1}$ (left-circular shift by k). The model can be visualized as a pointer on a dial with seven positions, as shown in Figure 3. Suppose that the errors consist of applying s_k with probability qe^{-k^2} , where $q = 0.5641$ is chosen so that the probabilities sum to 1, that is $\sum_{k=-\infty}^{\infty} qe^{-k^2} = 1$. Thus, s_0 has probability 0.5641, and each of s_{-1} and s_1 has probability 0.2075. These are the main errors that we need to protect against. Continuous versions of this error model in the context of communication channels are known as Gaussian channels.

One bit can be encoded in this physical system by the map $0 \rightarrow 1, 1 \rightarrow 4$. To decode with protection against $s_0, s_{-1},$ and s_1 , use the mapping

- 0 \rightarrow 0
- 1 \rightarrow 0
- 2 \rightarrow 0
- 3 \rightarrow 1
- 4 \rightarrow 1
- 5 \rightarrow 1
- 6 \rightarrow fail

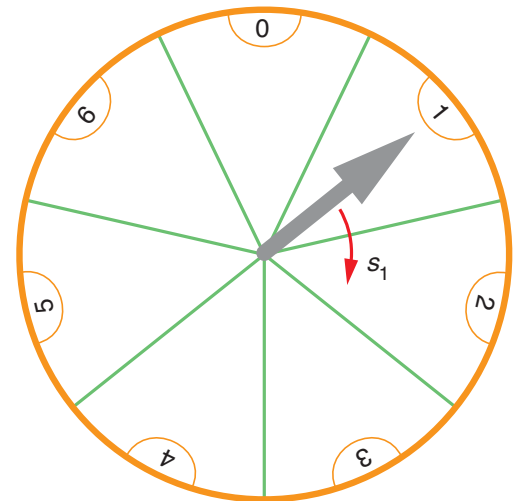


Figure 3. A Seven-State Cyclic System

- (4) The position of the pointer on the seven-position dial determines the state of the system. With the pointer in the position shown, the state is 1. Errors have the effect of rotating the pointer clockwise or counterclockwise. The effect of s_1 is to rotate the pointer clockwise, as shown by the red arrow.

(5)

If state 6 is encountered, we know that an error involving a shift of at least 2 (left or right) occurred, but there is no reasonable way of decoding it to the state of a bit. This means that the error is detected, but we cannot correct it. Error detection can be used by the receiver to ask for information to be sent again. The probability of correctly decoding with this code is at least 0.9792, which is the probability that the error caused a shift of at most 1.

As before, a pair of syndrome and information-carrying subsystems can be identified as being used by the encoding and decoding procedures. It suffices to correctly identify the syndrome states, which we name -1 , 0 , and 1 , because they indicate which of the likeliest shifts happened. The resulting subsystem identification is

$$\begin{array}{ll}
 0 & \leftrightarrow -1 \cdot 0 \\
 1 & \leftrightarrow 0 \cdot 0 \\
 2 & \leftrightarrow 1 \cdot 0 \\
 3 & \leftrightarrow -1 \cdot 1 \\
 4 & \leftrightarrow 0 \cdot 1 \\
 5 & \leftrightarrow 1 \cdot 1
 \end{array} \tag{6}$$

A new feature of this subsystem identification is that it is incomplete: Only a subset of the state space is identified. In this case, the complement can be used for error detection.

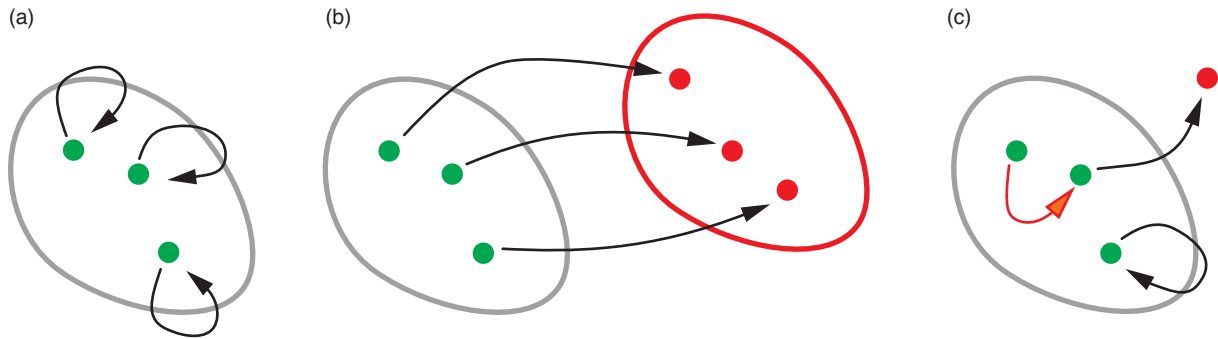
Like the repetition code, this code can be used in a setting where the errors happen repeatedly. Again, it suffices to reset the syndrome subsystem, in this case to 0 , to keep the encoded information protected. After the syndrome subsystem has been reset, a subsequent s_1 or s_{-1} error affects only the syndrome.

Principles of Error Correction

When considering the problem of limiting the effects of errors in information processing, the first task is to establish the properties of the physical systems that are available for representing and computing with information. Thus, it is necessary to learn the following: the physical system to be used, in particular the structure of its state space; the available means for controlling this system; the type of information to be processed; and the nature of the errors, that is, the error model. With this information, the approaches used to correct errors in the three examples provided in the previous section involve the following:

1. Determine a code, which is a subspace of the physical system, that can represent the information to be processed.
2. (a) Identify a decoding procedure that can restore the information represented in the code after any one of the most likely errors occurred or (b) determine a pair of syndrome and information-carrying subsystems such that the code corresponds to a “base” state of the syndrome subsystem and the primary errors act only on the syndrome.
3. Analyze the error behavior of the code and subsystem.

The tasks of determining a code and identifying decoding procedures or subsystems are closely related. As a result, the following questions are at the foundation of the theory of error correction: What properties must a code satisfy so that it can be used to protect well against a given error model? How does one obtain the decoding or subsystem identification that achieves this protection? In many cases, the answers can be based on choosing a fixed set of error operators that represents well the most likely errors and then determining whether these errors can be protected against without any loss of information. Once an error set is fixed, determining whether it is correctable can be cast in terms of the idea of detectable errors. This idea works equally well for both classical and quantum information. We introduce it using classical information concepts.



Error Detection. Error detection was used in the cyclic-system example to reject a state that could not be properly decoded. In the communication setting, error control methods based on error detection alone work as follows: The encoded information is transmitted. The receiver checks whether the state is still in the code, that is, whether it could have been obtained by encoding. If not, the result is rejected. The sender can be informed of the failure so that the information can be retransmitted. Given a set of error operators that need to be protected against, the scheme is successful if, for each error operator, either the information is unchanged or the error is detected. Thus, we can say that an operator E is detectable by a code if, for each state x in the code, either $Ex = x$ or Ex is not in the code (see Figure 4).

What errors are detectable by the codes in the examples? The code in the first example consists of 00 and 01 . Every operator that affects only the first bit is therefore detectable. In particular, all the operators in the error model are detectable. In the second example, the code consists of the states 000 and 111 . The identity operator has no effect and is therefore detectable. Any flips of exactly one or two bits are detectable because the states in the code are changed to states outside the code. The error that flips all bits is not detectable because it preserves the code but changes the states in the code. With the code for the cyclic system, shifts by -2 , -1 , 0 , 1 , and 2 are detectable but not shifts by 3 .

To conclude the section, we state a characterization of detectability, which has a natural generalization to the case of quantum information.

Theorem 1. E is detectable by a code if and only if for all $x \neq y$ in the code, $Ex \neq y$.

From Error Detection to Error Correction. Given a code C and a set of error operators $\mathcal{E} = \{1 = E_0, E_1, E_2, \dots\}$, is it possible to determine whether a decoding procedure or subsystem exists such that \mathcal{E} is correctable (by C), that is, such that the errors in \mathcal{E} do not affect the encoded information? As explained below, the answer is yes, and the solution is to check the condition in the following theorem:

Theorem 2. \mathcal{E} is correctable by C if and only if, for all $x \neq y$ in the code and all i and j , it is true that $E_i x \neq E_j y$.

Observe that the notion of correctability depends on all the errors in the set under consideration and, unlike detectability, cannot be applied to individual errors.

To see that the condition for correctability in Theorem 2 is necessary, suppose that for some $x \neq y$ in the code and some i and j , we have $z = E_i x = E_j y$. If the state z is obtained after an unknown error in \mathcal{E} , then it is not possible to determine whether the original code word was x or y because we cannot tell whether E_i or E_j occurred.

To see that the condition for correctability in Theorem 2 is sufficient, we assume it and construct a decoding method $z \rightarrow \text{dec}(z)$. Suppose that after an unknown error

Figure 4. Typical Detectable and Undetectable Code Errors Three examples are shown. In each, the code is represented by a brown oval containing three code words (green points). The effect of the error operator is shown as arrows. (a) The error does not change the code words and is therefore considered detectable. (b) The error maps the code words outside the code so that it is detected. (c) One code word is mapped to another, as shown by the red arrow. Finding that a received word is still in the code does not guarantee that it was the originally encoded word. The error is therefore not detectable.

occurred, the state z is obtained. There can be one and only one x in the code for which some $E_{i(z)} \in \mathcal{E}$ satisfies the condition that $E_{i(z)}x = z$. Thus, x must be the original code word, and we can decode z by defining $x = \text{dec}(z)$. Note that it is possible for two errors to have the same effect on some code words. A subsystem identification for this decoding is given by $z \leftrightarrow i(z) \cdot \text{dec}(z)$, where the syndrome subsystem's state space consists of error operator indices $i(z)$ and the information-carrying system's consists of the code words $\text{dec}(z)$ returned by the decoding. The subsystem identification thus constructed is not necessarily onto the state space of the subsystem pair. That is, for different code words x , the set of $i(z)$ such that $\text{dec}(z) = x$ can vary and need not be all the error indices. As we will show, the subsystem identification is onto the state space of the subsystem pair in the case of quantum information. It is instructive to check that, when applied to the examples, this subsystem construction does give a version of the subsystem identifications provided earlier.

It is possible to relate the condition for correctability of an error set to detectability. For simplicity, assume that each E_i is invertible. (This assumption is satisfied by our examples but not by error operators such as “reset bit one to 0.”) In this case, the correctability condition is equivalent to the statement that all products $E_j^{-1} E_i$ are detectable. To see the equivalence, first suppose that some $E_j^{-1} E_i$ is not detectable. Then, there are $x \neq y$ in the code such that $E_j^{-1} E_i x = y$. Consequently, $E_i x = E_j y$, and the error set is not correctable. This argument can be reversed to complete the proof of equivalence.

If the assumption that the errors are invertible does not hold, the relationship between detectability and correctability becomes more complicated, requiring a generalization of the inverse operation. This generalization is simpler in the quantum setting.

Quantum Error Correction

The principles of error correction outlined before apply to the quantum setting as readily as to the classical setting. The main difference is that the physical system to be used for representing and processing information behaves quantum mechanically and the type of information is quantum. The question of how classical information can be protected in quantum systems is also interesting but will not be discussed here. We illustrate the principles of quantum error correction by considering quantum versions of the three examples given in “Concepts and Examples” and then add a uniquely quantum example with potentially practical applications in, for example, quantum dot technologies. For an explanation of the basic quantum-information concepts and conventions, see the article “Quantum Information Processing” on page 2.

Trivial Two-Qubit Example. A quantum version of the two-bit example from the previous section consists of two physical qubits, where the errors randomly apply the identity or one of the Pauli operators to the first qubit. The Pauli operators are defined by

$$\mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \text{ and } \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (7)$$

Explicitly, the errors have the effect

$$|\psi\rangle_{12} \begin{cases} \mathbb{1}|\psi\rangle_{12} & \text{Probability } .25 \\ \sigma_x^{(1)}|\psi\rangle_{12} & \text{Probability } .25 \\ \sigma_y^{(1)}|\psi\rangle_{12} & \text{Probability } .25 \\ \sigma_z^{(1)}|\psi\rangle_{12} & \text{Probability } .25 \end{cases}, \quad (8)$$

where the superscripts in parentheses specify the qubit that an operator acts on. This error model is called completely depolarizing on qubit 1. Obviously, a one-qubit state can be stored in the second physical qubit without being affected by the errors. An encoding operation that implements this observation is

$$|\psi\rangle \rightarrow |\circ\rangle_1 |\psi\rangle_2, \quad (9)$$

which realizes an ideal qubit as a two-dimensional subspace of the physical qubits. This subspace is the quantum code for this encoding. To decode, one can discard physical qubit 1 and return qubit 2, which is considered a natural subsystem of the physical system. In this case, the identification of syndrome and information-carrying subsystems is the obvious one associated with the two physical qubits.

Quantum Repetition Code. The repetition code can be used to protect quantum information in the presence of a restricted error model. Let the physical system consist of three qubits. Errors act by independently applying, to each qubit, the flip operator σ_x with probability .25. The classical code can be made into a quantum code by the superposition principle. Encoding one qubit is accomplished by

$$\alpha|\circ\rangle + \beta|1\rangle \rightarrow \alpha|\circ\circ\circ\rangle + \beta|111\rangle. \quad (10)$$

The associated quantum code is the range of the encoding, that is, the two-dimensional subspace spanned by the encoded states $|\circ\circ\circ\rangle$ and $|111\rangle$.

As in the classical case, decoding is accomplished by majority logic. However, it must be implemented carefully to avoid destroying quantum coherence in the stored information. One way to do that is to use only unitary operations to transfer the stored information to the output qubit. Figure 5 shows a quantum network that accomplishes this task.

As shown, the decoding network establishes an identification between the three physical qubits and a pair of subsystems consisting of two qubits representing the syndrome subsystem and one qubit for the information-carrying subsystem. On the left side of the correspondence, the information-carrying subsystem is not identifiable with any one (or two) of the physical qubits. Nevertheless, it exists there through the identification.

To obtain a network for encoding, we reverse the decoding network and initialize qubits 2 and 3 in the state $|\circ\circ\rangle$. The initialization renders the Toffoli gate unnecessary. The complete system with a typical error is shown in Figure 6.

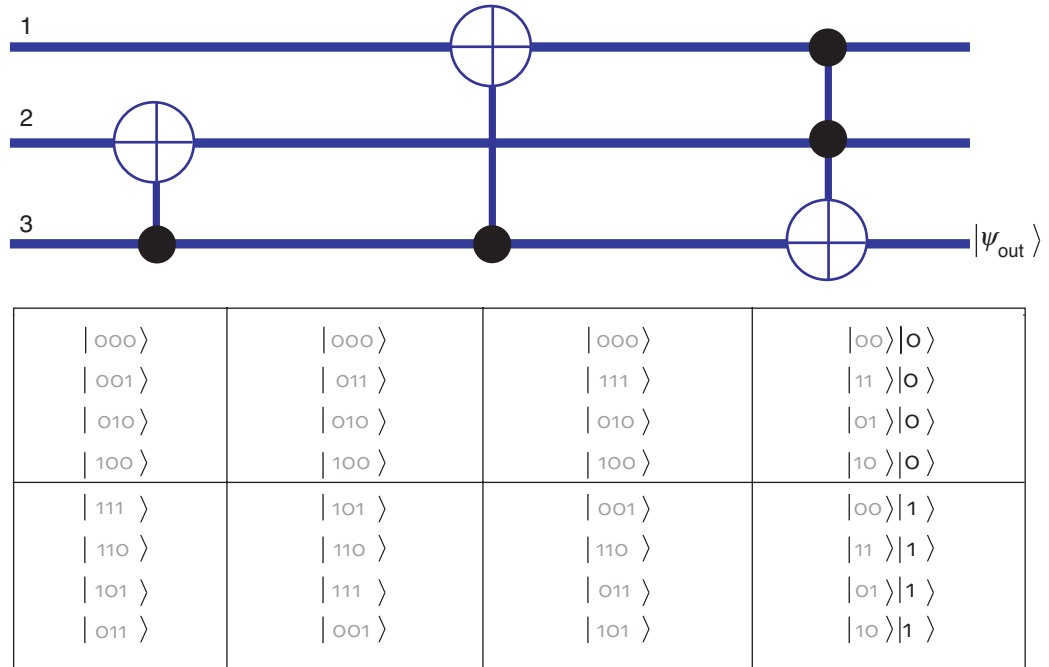


Figure 5. Majority Logic Decoding into the Output Qubit 3

The effect of the quantum network on the basis states is shown. The top half shows the states with majority 0. The decoded qubit is separated in the last step. The conventions for illustrating quantum networks are explained in the article “Quantum Information Processing” on page 2.

As in the case of the classical repetition code, we can protect against cumulative errors without explicitly decoding and then reencoding, which would cause a temporary loss of protection. Instead, one can find a means for directly resetting the syndrome subsystem to $|00\rangle$ (thus returning the information to the code) before the errors happen again. After resetting in this way, the errors in the correctable set have no effect on the encoded information because they act only on the syndrome subsystem.

Part of the task of designing error-correcting systems is to determine how well the system performs. An important performance measure is the probability of error. In quantum systems, the probability of error is intuitively interpreted as the maximum probability with which we can see a result different from the expected one in any measurement. Specifically, to determine the error, one compares the output $|\psi_o\rangle$ of the system with the input $|\psi\rangle$. An upper bound is obtained if the output is written as a combination of the input state and an error state. For quantum information, combinations are linear combinations (that is, superpositions). Thus $|\psi_o\rangle = \gamma |\psi\rangle + |e\rangle$ (see Figure 7). The probability of error is bounded by $\epsilon = ||e||^2$ (which we call an error estimate). In general, there are many different ways of writing the output as a combination of an acceptable state and an error term. One attempts to choose the combination that minimizes the error estimate. This choice yields the number ϵ for which $1 - \epsilon$ is called fidelity. A fidelity of 1 means that the output is the same (up to a phase factor) as the input.

To illustrate error analysis, we calculate the error for the repetition code example for the two initial states $|0\rangle$ and $(1/\sqrt{2})(|0\rangle + |1\rangle)$.

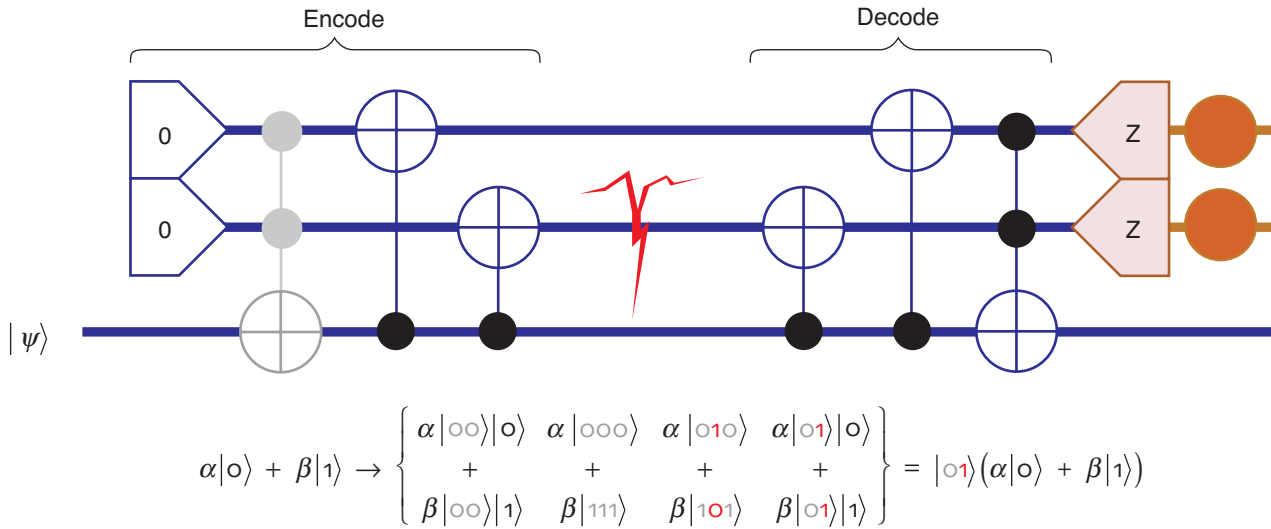


Figure 6. Networks for the Quantum Repetition Code with a Typical Error
 The error that occurred can be determined from the state of the syndrome subsystem, which consists of the top two qubits. The encoding is shown as the reverse of the decoding, starting with an initialized syndrome subsystem. When the decoding is reversed to yield the encoding, there is an initial Toffoli gate (shown in gray). Because of the initialization, this gate has no effect and is therefore omitted in an implementation.

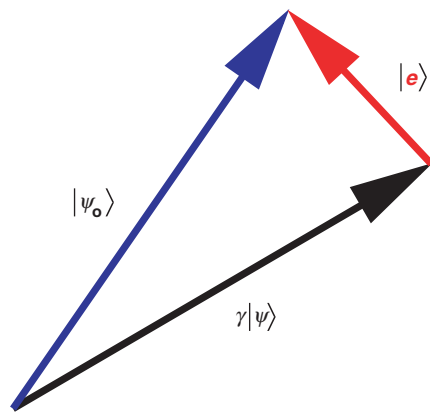


Figure 7. Error Estimate
 Any decomposition of the output state $|\psi_0\rangle$ into a “good” state $\gamma|\psi\rangle$ and an (unnormalized) error term $|e\rangle$ gives an estimate $\epsilon = \|e\|^2$. For pure states, the optimum estimate is obtained when the error term is orthogonal to the input state. To obtain an error estimate for mixtures, one can use any representation of the state as a probabilistic combination of pure states and calculate the probabilistic sum of the pure-state errors.

$$|0\rangle \xrightarrow{\text{encode}} |000\rangle \tag{11}$$

$$\xrightarrow{\text{error}} \begin{cases} .75^3 : |000\rangle , \\ .25 * .75^2 : |100\rangle , \\ .25 * .75^2 : |010\rangle , \\ .25 * .75^2 : |001\rangle , \\ .25^2 * .75 : |110\rangle , \\ .25^2 * .75 : |101\rangle , \\ .25^2 * .75 : |011\rangle , \\ .25^3 : |111\rangle . \end{cases} \tag{12}$$

$$\xrightarrow{\text{decode}} \begin{cases} .4219 : |00\rangle \cdot |0\rangle , \\ .1406 : |10\rangle \cdot |0\rangle , \\ .1406 : |01\rangle \cdot |0\rangle , \\ .1406 : |11\rangle \cdot |0\rangle , \\ .0469 : |11\rangle \cdot |1\rangle , \\ .0469 : |01\rangle \cdot |1\rangle , \\ .0469 : |10\rangle \cdot |1\rangle , \\ .0156 : |00\rangle \cdot |1\rangle . \end{cases} \tag{13}$$

The final state is a mixture consisting of four correctly decoded components and four incorrectly decoded ones. The probability of each state in the mixture is shown before the colon. The incorrectly decoded information is orthogonal to the encoded information, and its probability is 0.1563, an improvement over the one-qubit error probability of 0.25. The second state behaves quite differently:

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \xrightarrow{\text{encode}} \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle) \tag{14}$$

$$\xrightarrow{\text{error}} \begin{cases} \vdots \\ .25^2 * .75 : \frac{1}{\sqrt{2}} (|110\rangle + |001\rangle) \\ \vdots \end{cases} \tag{15}$$

$$\xrightarrow{\text{decode}} \begin{cases} \vdots \\ .0469 : \frac{1}{\sqrt{2}} |11\rangle \cdot (|1\rangle + |0\rangle) \\ \vdots \end{cases} . \tag{16}$$

Not all error events have been shown, but in each case it can be seen that the state is decoded correctly, so the error is 0. This shows that the error probability can depend

significantly on the initial state. To remove this dependence and give a state independent error quantity, one can use the worst-case, the average, or the entanglement error. See the section “Quantum Error Analysis” on page 209.

Quantum Code for a Cyclic System. The shift operators introduced earlier act as permutations of the seven states of the cyclic system. They can therefore be extended to unitary operators on a seven-state cyclic quantum system with logical basis $|0\rangle, |1\rangle, |2\rangle, |3\rangle, |4\rangle, |5\rangle,$ and $|6\rangle$. The error model introduced earlier makes sense here without modification, as does the encoding. The subsystem identification now takes the six-dimensional subspace spanned by $|0\rangle, \dots, |5\rangle$ to a pair consisting of a three-state system with basis $|{-1}\rangle, |0\rangle, |1\rangle$ and a qubit. The identification of Equation (6) extends linearly to a unitary subsystem identification. The procedure for decoding is modified as follows: First, a measurement determines whether the state is in the six-dimensional subspace or not. If it is, the identification is used to extract the qubit. Here is an outline of what happens when the state $(1/\sqrt{2})(|0\rangle + |1\rangle)$ is encoded:

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \xrightarrow{\text{encode}} \frac{1}{\sqrt{2}}(|1\rangle + |4\rangle) \tag{17}$$

$$\xrightarrow{\text{red arrow}} \left\{ \begin{array}{l} \vdots \\ .05641e^{-4} \vdots \frac{1}{\sqrt{2}}(|3\rangle + |6\rangle) \\ \vdots \end{array} \right. \tag{18}$$

$$\xrightarrow{\text{detect}} \left\{ \begin{array}{l} \vdots \\ .001 : \left\{ \begin{array}{l} .5 : \text{fail} \\ \vdots \\ .5 : |3\rangle \end{array} \right. \\ \vdots \end{array} \right. \tag{19}$$

$$\xrightarrow{\text{decode}} \left\{ \begin{array}{l} \vdots \\ .0005 : \text{fail} \\ .0005 : |{-1}\rangle \cdot |1\rangle \\ \vdots \end{array} \right. \tag{20}$$

$$= \left\{ \begin{array}{l} \vdots \\ .0005 : \text{fail} \\ .0005 : |{-1}\rangle \cdot \left(\frac{1}{2}(|0\rangle + |1\rangle) + \frac{1}{2}(|{-0}\rangle + |1\rangle) \right) \\ \vdots \end{array} \right. \tag{21}$$

A “good” state was separated from the output in the case that is shown. The leftover error term has probability amplitude $.0005 * ((1/2)^2 + (1/2)^2) = .00025$, which contributes to the total error (not shown).

Three Quantum Spin-1/2 Particles. Quantum physics provides a rich source of systems with many opportunities for representing and protecting quantum information. Sometimes, it is possible to encode information in such a way that it is protected from the errors indefinitely, without intervention. An example is the trivial two-qubit system discussed before. Whenever error protection without intervention is possible, there is an information-carrying subsystem such that errors act only on the associated syndrome subsystem regardless of the current state. An information-carrying subsystem with this property is called “noiseless.” A physically motivated example of a one-qubit noiseless

subsystem can be found in three spin-1/2 particles with errors due to random fluctuations in an external field.

A spin-1/2 particle's state space is spanned by two states, $|\uparrow\rangle$ and $|\downarrow\rangle$. Intuitively, these states correspond to the spin pointing “up” ($|\uparrow\rangle$) or “down” ($|\downarrow\rangle$) in some chosen reference frame. The state space is therefore the same as that of a qubit, and we can make the identifications $|\uparrow\rangle \leftrightarrow |0\rangle$ and $|\downarrow\rangle \leftrightarrow |1\rangle$. An external field causes the spin to rotate according to an evolution of the form

$$|\psi_t\rangle = e^{-i(u_x\sigma_x + u_y\sigma_y + u_z\sigma_z)t/2} |\psi\rangle. \quad (22)$$

The vector $\mathbf{u} = (u_x, u_y, u_z)$ characterizes the direction of the field and the strength of the spin's interaction with the field. This situation arises, for example, in nuclear magnetic resonance with spin-1/2 nuclei, where the fields are magnetic fields (see the article “NMR and Quantum Information Processing” on page 226).

Now consider the physical system composed of three spin-1/2 particles with errors acting as identical rotations of the three particles. Such errors occur if they are due to a uniform external field that fluctuates randomly in direction and strength. The evolution caused by a uniform field is given by

$$\begin{aligned} |\psi_t\rangle_{123} &= e^{-i(u_x\sigma_x^{(1)} + u_y\sigma_y^{(1)} + u_z\sigma_z^{(1)})t/2} e^{-i(u_x\sigma_x^{(2)} + u_y\sigma_y^{(2)} + u_z\sigma_z^{(2)})t/2} e^{-i(u_x\sigma_x^{(3)} + u_y\sigma_y^{(3)} + u_z\sigma_z^{(3)})t/2} |\psi\rangle_{123} \\ &= e^{-i(u_x(\sigma_x^{(1)} + \sigma_x^{(2)} + \sigma_x^{(3)}) + u_y(\sigma_y^{(1)} + \sigma_y^{(2)} + \sigma_y^{(3)}) + u_z(\sigma_z^{(1)} + \sigma_z^{(2)} + \sigma_z^{(3)}))t/2} |\psi\rangle_{123} \\ &= e^{-i(u_x J_x + u_y J_y + u_z J_z)t} |\psi\rangle_{123}, \end{aligned} \quad (23)$$

with $J_u = (\sigma_u^{(1)} + \sigma_u^{(2)} + \sigma_u^{(3)})/2$ for $u = x, y, \text{ and } z$. We can exhibit the error operators arising from a uniform field in a compact form by defining $\mathbf{J} = (J_x, J_y, J_z)$ and $\mathbf{v} = (u_x, u_y, u_z)t$. Then the error operators are given by $E(\mathbf{v}) = e^{-i\mathbf{v}\cdot\mathbf{J}}$, where the dot product in the exponent is calculated like the standard vector dot product.

For a one-qubit noiseless subsystem, the key property of the error model is that the errors are symmetric under any permutation of the three particles. A permutation of the particles acts on the particles' state space by permuting the labels in the logical states. For example, the permutation π that swaps the first two particles acts on logical states as

$$\pi|a\rangle_1|b\rangle_2|c\rangle_3 = |a\rangle_2|b\rangle_1|c\rangle_3 = |b\rangle_1|a\rangle_2|c\rangle_3. \quad (24)$$

To say that the errors are symmetric under particle permutations means that each error E satisfies $\pi^{-1}E\pi = E$, or equivalently, $E\pi = \pi E$ (E commutes with π). To see that this condition is satisfied, write

$$\begin{aligned}
\pi^{-1}E(\mathbf{v})\pi &= \pi^{-1}e^{-i\mathbf{v}\cdot\mathbf{J}}\pi \\
&= e^{-i\pi^{-1}(\mathbf{v}\cdot\mathbf{J})\pi} \\
&= e^{-i\mathbf{v}\left(\pi^{-1}\mathbf{J}\pi\right)}.
\end{aligned} \tag{25}$$

If π permutes particle a to particle b , then $\pi^{-1}\sigma_u^{(a)}\pi = \sigma_u^{(b)}$. It follows that $\pi^{-1}\mathbf{J}\pi = \mathbf{J}$. This expression shows that the errors commute with the particle permutations and therefore cannot distinguish between the particles. An error model satisfying this property is called a collective error model.

If a noiseless subsystem exists, then learning the symmetries of the error model suffices for constructing the subsystem. This procedure is explained later, in ‘‘Conserved Quantities, Symmetries, and Noiseless Subsystems.’’ For the three spin-1/2 system, the procedure results in a one-qubit noiseless subsystem protected from all collective errors. We first exhibit the subsystem identification and then discuss its properties to explain why it is noiseless. As in the case of the seven-state cyclic system, the identification involves a proper subspace of the physical system’s state space. The subsystem identification involves a four-dimensional subspace and is defined by the following correspondence:

$$\begin{aligned}
&\frac{1}{\sqrt{3}}\left(|\downarrow\rangle_1|\uparrow\rangle_2|\uparrow\rangle_3 + e^{-i2\pi/3}|\uparrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3 + e^{i2\pi/3}|\uparrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3\right) \leftrightarrow |\uparrow\rangle \cdot |\circ\rangle \\
&\frac{1}{\sqrt{3}}\left(|\downarrow\rangle_1|\uparrow\rangle_2|\uparrow\rangle_3 + e^{i2\pi/3}|\uparrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3 + e^{-i2\pi/3}|\uparrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3\right) \leftrightarrow |\uparrow\rangle \cdot |\uparrow\rangle \\
&-\frac{1}{\sqrt{3}}\left(|\uparrow\rangle_1|\downarrow\rangle_2|\downarrow\rangle_3 + e^{-i2\pi/3}|\downarrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3 + e^{i2\pi/3}|\downarrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3\right) \leftrightarrow |\downarrow\rangle \cdot |\circ\rangle \\
&-\frac{1}{\sqrt{3}}\left(|\uparrow\rangle_1|\downarrow\rangle_2|\downarrow\rangle_3 + e^{i2\pi/3}|\downarrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3 + e^{-i2\pi/3}|\downarrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3\right) \leftrightarrow |\downarrow\rangle \cdot |\uparrow\rangle.
\end{aligned} \tag{26}$$

The state labels for the syndrome subsystem (before the dot in the expressions on the right side) identify it as a spin-1/2 subsystem. In particular, it responds to the errors caused by uniform fields in the same way as the physical spin-1/2 particles. This behavior is caused by $2J_u$ acting as the u -Pauli operator on the syndrome subsystem.

To confirm this property, we apply $2J_u$ to the logical states of Equation (26) for $u = z, x$. The property for $u = y$ then follows because $i\sigma_y = \sigma_z\sigma_x$. Consider $2J_z$. Each of the four states shown in Equation (26) is an eigenstate of $2J_z$. For example, the physical state for $|\uparrow\rangle \cdot |\circ\rangle$ is a superposition of states with two spins up (\uparrow) and one spin down (\downarrow). The eigenvalue of such a state with respect to $2J_z$ is the difference Δ between the number of spins that are up and down. Thus, $2J_z|\uparrow\rangle \cdot |\circ\rangle = |\uparrow\rangle \cdot |\circ\rangle$. The difference is also $\Delta = 1$ for $|\uparrow\rangle \cdot |\uparrow\rangle$ and $\Delta = -1$ for $|\downarrow\rangle \cdot |\circ\rangle$ and $|\downarrow\rangle \cdot |\uparrow\rangle$. Therefore, $2J_z$ acts as the z -Pauli operator on the syndrome subsystem. To confirm this behavior for $2J_x$, we compute $2J_x|\uparrow\rangle \cdot |\circ\rangle$.

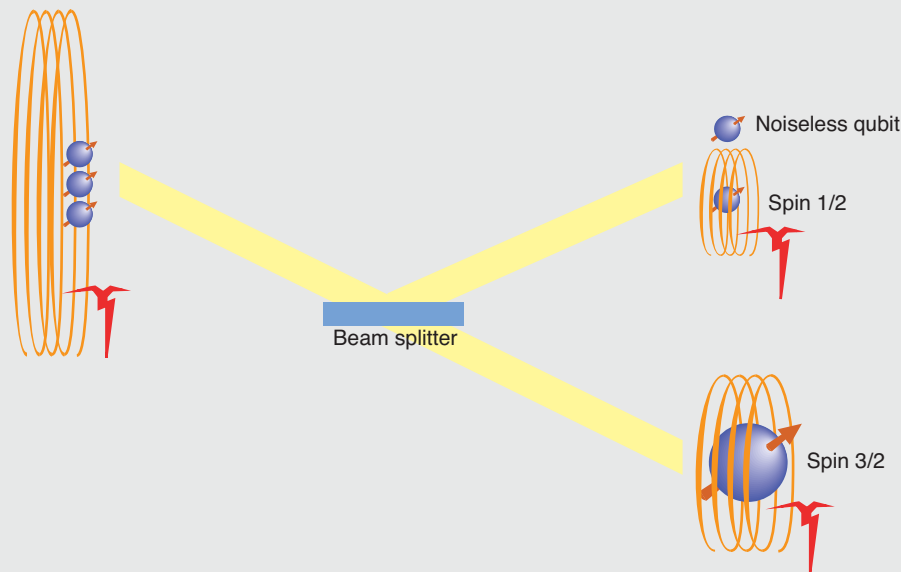
$$\begin{aligned}
 2J_x|\uparrow\rangle\cdot|\circ\rangle &= 2J_x\frac{1}{\sqrt{3}}\left(|\downarrow\rangle_1|\uparrow\rangle_2|\uparrow\rangle_3 + e^{-i2\pi/3}|\uparrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3 + e^{i2\pi/3}|\uparrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3\right) \\
 &= \frac{1}{\sqrt{3}}\left(\sigma_x^{(1)} + \sigma_x^{(2)} + \sigma_x^{(3)}\right)|\downarrow\rangle_1|\uparrow\rangle_2|\uparrow\rangle_3 \\
 &\quad + e^{-i2\pi/3}\frac{1}{\sqrt{3}}\left(\sigma_x^{(1)} + \sigma_x^{(2)} + \sigma_x^{(3)}\right)|\uparrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3 \\
 &\quad + e^{i2\pi/3}\frac{1}{\sqrt{3}}\left(\sigma_x^{(1)} + \sigma_x^{(2)} + \sigma_x^{(3)}\right)|\uparrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3 \\
 &= \frac{1}{\sqrt{3}}\left(|\uparrow\rangle_1|\uparrow\rangle_2|\uparrow\rangle_3 + |\downarrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3 + |\downarrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3\right) \\
 &\quad + e^{-i2\pi/3}\frac{1}{\sqrt{3}}\left(|\downarrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3 + |\uparrow\rangle_1|\uparrow\rangle_2|\uparrow\rangle_3 + |\uparrow\rangle_1|\downarrow\rangle_2|\downarrow\rangle_3\right) \\
 &\quad + e^{i2\pi/3}\frac{1}{\sqrt{3}}\left(|\downarrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3 + |\uparrow\rangle_1|\downarrow\rangle_2|\downarrow\rangle_3 + |\uparrow\rangle_1|\uparrow\rangle_2|\uparrow\rangle_3\right) \\
 &= \frac{1}{\sqrt{3}}\left(1 + e^{-i2\pi/3} + e^{i2\pi/3}\right)|\uparrow\rangle_1|\uparrow\rangle_2|\uparrow\rangle_3 \\
 &\quad + \frac{1}{\sqrt{3}}\left(e^{-i2\pi/3} + e^{i2\pi/3}\right)|\uparrow\rangle_1|\downarrow\rangle_2|\downarrow\rangle_3 \\
 &\quad + \frac{1}{\sqrt{3}}\left(1 + e^{i2\pi/3}\right)|\downarrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3 \\
 &\quad + \frac{1}{\sqrt{3}}\left(1 + e^{-i2\pi/3}\right)|\downarrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3 \\
 &= -\frac{1}{\sqrt{3}}\left(|\uparrow\rangle_1|\downarrow\rangle_2|\downarrow\rangle_3 + e^{-i2\pi/3}|\downarrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3 + e^{i2\pi/3}|\downarrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3\right) \\
 &= |\downarrow\rangle\cdot|\circ\rangle . \tag{27}
 \end{aligned}$$

Similarly, one can check that, for the other logical states, the effect of $2J_x$ is to flip the orientation of the syndrome spin. That the subsystem identified in Equation (26) is noiseless now follows from the fact that the errors $E(\mathbf{v})$ are exponentials of sums of the syndrome spin operators J_u . The errors therefore act as the identity on the information-carrying subsystem.

The noiseless qubit supported by three spin-1/2 particles with collective errors is another example in which the subsystem identification does not involve the whole state space of the system. In this case, the errors of the error model cannot remove amplitude from the subspace. As a result, if we detect an error, that is, if we find that the system's state is in the orthogonal complement of the subspace of the subsystem identification, we can deduce that either the error model is inadequate or we introduced errors in the manipulations required for transferring information to the noiseless qubit.

The noiseless subsystem of three spin-1/2 particles can be physically motivated by an analysis of quantum spin numbers. This analysis is outlined in the box on the opposite page.

Creating a Noiseless Subsystem from Three Spin-1/2 Particles



The left side shows the three particles with errors caused by fluctuations in a uniform magnetic field depicted by a noisy coil. The spin along direction u ($u = x, y, z$) can be measured, and its expectation is given by $\langle \psi | J_u | \psi \rangle$, where $|\psi\rangle$ is the quantum state of the particles and J_u is the total spin observable along the u -axis given by the half sum of the u -Pauli matrices of the particles as defined in the text. The squared magnitude of the total spin is given by the expectation of the observable $J^2 = \mathbf{J} \cdot \mathbf{J} = J_x^2 + J_y^2 + J_z^2$. The observable J^2 commutes with the J_u and therefore also with the errors $E(\mathbf{v}) = e^{-i\mathbf{v} \cdot \mathbf{J}}$ caused by uniform field fluctuations. This statement can be verified directly, or one can note that $E(\mathbf{v})$ acts on \mathbf{J} as a rotation in three dimensions, and as one would expect, such rotations preserve the squared length J^2 of \mathbf{J} . It now follows that the eigenspaces of J^2 are invariant under the errors and, therefore, that the eigenspaces are good places to look for noiseless subsystems. The eigenvalues of J^2 are of the form $j(j+1)$, where j is the spin quantum number of the corresponding eigenspace. There are two eigenspaces, one with spin $j = 1/2$ and the other with spin $j = 3/2$.

The figure shows a thought experiment that involves passing the three-particle system through a type of beam splitter or Stern-Gerlach apparatus sensitive to J^2 . Using such a beam splitter, the system of particles can be made to go in one of two directions, depend-

ing on j . In the figure, if the system's state is in the spin-3/2 subspace, it passes through the beam splitter; if it is in the spin-1/2 subspace, the system is reflected up. It can be shown that the subspace with $j = 3/2$ is four dimensional and spanned by the states that are symmetric under particle permutations. Unfortunately, there is no noiseless subsystem in this subspace (refer to the section "Conserved Quantities, Symmetries, and Noiseless Subsystems"). The spin-1/2 subspace is also four dimensional and spanned by the states in Equation (26). The spin-1/2 property of the subspace implies that the spin operators J_u act in a way that is algebraically identical to the way $\sigma_u/2$ acts on a single spin-1/2 particle. This property implies the existence of the syndrome subsystem introduced in the text. Conventionally, the spin-1/2 subspace is thought of as consisting of two orthogonal two-dimensional subspaces, each behaving like a spin-1/2 with respect to the J_u . This choice of subspaces is not unique, but by associating them with two logical states of a noiseless qubit, one can obtain the subsystem identification of Equation (26). Some care needs to be taken to ensure that the noiseless qubit operators commute with the J_u , as they should. In the thought experiment shown in the figure, one can imagine unitarily rotating the system emerging in the upper path to make explicit the syndrome spin-1/2 subsystem and the noiseless qubit with which it must be paired. The result of this rotation is shown.

Error Models

We have seen several models of physical systems and errors in the examples of the previous sections. Most physical systems under consideration for QIP consist of particles or degrees of freedom that are spatially localized, a feature reflected in the error models that are usually investigated. Because we also expect the physically realized qubits to be localized, the standard error models deal with quantum errors that act independently on different qubits. Logically realized qubits, such as those implemented by subsystems different from the physically obvious ones, may have more complicated residual-error behaviors.

The Standard Error Models for Qubits. The most investigated error model for qubits consists of independent, depolarizing errors. This model has the effect of completely depolarizing each qubit independently with probability p —see Equation (8). For one qubit, the model is the least biased in the sense that it is symmetric under rotations. As a result, every state of the qubit is equally affected. Independent depolarizing errors are considered to be the quantum analogue of the classical independent bit-flip error model.

Depolarizing errors are not typical for physically realized qubits. However, given the ability to control individual qubits, it is possible to enforce the depolarizing model (see below). Consequently, error correction methods designed to control depolarizing errors apply to all independent error models. Nevertheless, it is worth keeping in mind that given detailed knowledge of the physical errors, a special purpose method is usually better than one designed for depolarizing errors. We therefore begin by showing how one can think about arbitrary error models.

There are several different ways of describing errors affecting a physical system (or “sys” for short) of interest. For most situations, in particular if the initial state of the system is pure, errors can be thought of as being the result of coupling to an initially independent environment for some time. Because of this coupling, the effect of error can always be represented by the process of adjoining an environment (or “env” for short) in some initial state $|0\rangle_{\text{env}}$ to the arbitrary state $|\psi\rangle_{\text{sys}}$, followed by a unitary coupling evolution $U^{(\text{env}, \text{sys})}$ acting jointly on the environment and the system. Symbolically, the process can be written as the map

$$|\psi\rangle_{\text{sys}} \rightarrow U^{(\text{env}, \text{sys})}|0\rangle_{\text{env}}|\psi\rangle_{\text{sys}} . \quad (28)$$

Choosing an arbitrary orthonormal basis consisting of the states $|e\rangle_{\text{env}}$ for the state space of the environment, the process can be rewritten in the form

$$\begin{aligned} |\psi\rangle_{\text{sys}} &\rightarrow \mathbb{1}^{(\text{env})} U^{(\text{env}, \text{sys})}|0\rangle_{\text{env}}|\psi\rangle_{\text{sys}} \\ &= \left(\sum_e |e\rangle_{\text{env}} \langle e| \right) U^{(\text{env}, \text{sys})}|0\rangle_{\text{env}}|\psi\rangle_{\text{sys}} \\ &= \sum_e |e\rangle_{\text{env}} \left(\langle e| U^{(\text{env}, \text{sys})}|0\rangle_{\text{env}} \right) |\psi\rangle_{\text{sys}} \\ &= \sum_e |e\rangle_{\text{env}} A_e^{(\text{sys})} |\psi\rangle_{\text{sys}} , \end{aligned} \quad (29)$$

where the last step defines operators $A_e^{(\text{sys})}$ acting on the physical system by $A_e^{(\text{sys})} = \text{env}\langle e|U^{(\text{env, sys})}|0\rangle_{\text{env}}$. The expression $\sum_e |e\rangle_{\text{env}} A_e^{(\text{sys})}$ is called an environment-labeled operator. The unitarity condition implies that $\sum_e A_e^\dagger A_e = \mathbb{1}$ (with system labels omitted). The environment basis $|e\rangle_{\text{env}}$ need not represent any physically meaningful choice of basis of a real environment. For error analysis, the states $|e\rangle_{\text{env}}$ are formal states that label the error operators A_e . One can use an expression of the form shown in Equation (29) even when the $|e\rangle$ are not normalized or orthogonal, keeping in mind that, as a result, the identity implied by the unitarity condition changes.

Note that the state on the right side of Equation (29), representing the effect of the errors, is correlated with the environment. This means that after removing (or “tracing over”) the environment, the state of the physical system is usually mixed. Instead of introducing an artificial environment, we can also describe the errors by using the density operator formalism for mixed states. Define $\rho = |\psi\rangle_{\text{sys}}\langle\psi|$. The effect of the errors on the density matrix ρ is given by the transformation

$$\rho \rightarrow \sum_e A_e \rho A_e^\dagger . \quad (30)$$

This is the “operator sum” formalism (Kraus 1983).

The two ways of writing the effects of errors can be applied to the depolarizing-error model for one qubit. As an environment-labeled operator, depolarization with probability p can be written as

$$\sqrt{1-p}|0\rangle_{\text{env}}\mathbb{1} + \frac{\sqrt{p}}{2} \left(|1\rangle_{\text{env}}\mathbb{1} + |x\rangle_{\text{env}}\sigma_x + |y\rangle_{\text{env}}\sigma_y + |z\rangle_{\text{env}}\sigma_z \right) , \quad (31)$$

where we introduced five abstract, orthonormal environment states to label the different events. In this case, one can think of the model as applying no error with probability $1-p$ or completely depolarizing the qubit with probability p . The latter event is represented by applying one of $\mathbb{1}$, σ_x , σ_y , or σ_z with equal probability $p/4$. To be able to think of the model as randomly applied Pauli matrices, it is crucial that the environment states labeling the different Pauli matrices be orthogonal. The square roots of the probabilities appear in the operator because, in an environment-labeled operator, it is necessary to give quantum amplitudes. Environment-labeled operators are useful primarily because of their great flexibility and redundancy.

In the operator sum formalism, depolarization with probability p transforms the input density matrix ρ as

$$\begin{aligned} \rho &\rightarrow (1-p)\rho + \frac{p}{4} (\mathbb{1}\rho\mathbb{1} + \sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z) \\ &= (1-3p/4)\rho + \frac{p}{4} (\sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z) . \end{aligned} \quad (32)$$

Because the operator sum formalism has less redundancy, it is easier to tell when two error effects are equivalent.

In the remainder of this section, we discuss how one can use active intervention to simplify the error model. To realize this simplification, we intentionally randomize the

qubit so that the environment cannot distinguish between the different axes defined by the Pauli spin matrices. Here is a simple randomization that actively converts an arbitrary error model for a qubit into one that consists of randomly applying Pauli operators according to some distribution. The distribution is not necessarily uniform, so the new error model is not yet depolarizing. Before the errors act, apply a random Pauli operator σ_u ($u = 0, x, y, z, \sigma_0 = 1$). After the errors act, apply the inverse of that operator, $\sigma_u^{-1} = \sigma_u$; then “forget” which operator was applied. This randomization method is called twirling (Bennett et al. 1996). To understand twirling, we use environment-labeled operators to demonstrate some of the techniques useful in this context. The sequence of actions implementing twirling can be written as follows (omitting labels for the physical system):

$$\begin{aligned}
 |\psi\rangle &\rightarrow \frac{1}{2} \sum_u |u\rangle_C \sigma_u |\psi\rangle && \text{Apply a random } \sigma_u \text{ remembering } u \text{ with the} \\
 &&& \text{help of the system } C. \\
 &\rightarrow \sum_e |e\rangle_{\text{env}} \frac{1}{2} \sum_u |u\rangle_C A_e \sigma_u |\psi\rangle && \text{Errors act.} \\
 &\rightarrow \sum_e |e\rangle_{\text{env}} \frac{1}{2} \sum_u |u\rangle_C \sigma_u A_e \sigma_u |\psi\rangle && \text{Apply } \sigma_u = \sigma_u^{-1}. \\
 &\rightarrow \sum_{eu} |eu\rangle_{\text{env},C} \frac{1}{2} \sigma_u A_e \sigma_u |\psi\rangle \cdot && \text{Forget which } u \text{ was used by absorbing} \quad (33) \\
 &&& \text{its memory in the environment.}
 \end{aligned}$$

The system C that was artificially introduced to carry the memory of u may be a classical memory because there is no need for coherence between different $|u\rangle_C$.

To determine the equivalent random Pauli operator error model, it is necessary to rewrite the total effect of the procedure using an environment-labeled sum involving orthogonal environment states and Pauli operators. To do so, express A_e as a sum of the Pauli operators, $A_e = \sum_v \alpha_{ev} \sigma_v$, using the fact that the σ_v are a linear basis for the space of one-qubit operators. Recall that σ_u anticommutes with σ_v if $0 \neq u \neq v \neq 0$. Thus, $\sigma_u \sigma_v \sigma_u = (-1)^{\langle v,u \rangle} \sigma_v$, where $\langle v, u \rangle = 1$ if $0 \neq u \neq v \neq 0$ and $\langle v, u \rangle = 0$ otherwise. We can now rewrite the last expression of Equation (33) as follows:

$$\begin{aligned}
 \sum_{eu} |eu\rangle_{\text{env},C} \frac{1}{2} \sigma_u A_e \sigma_u |\psi\rangle &= \sum_{eu} |eu\rangle_{\text{env},C} \frac{1}{2} \sigma_u \sum_v \alpha_{ev} \sigma_v \sigma_u |\psi\rangle \\
 &= \sum_v \left(\sum_{eu} \frac{1}{2} \alpha_{ev} (-1)^{\langle v,u \rangle} |eu\rangle_{\text{env},C} \right) \sigma_v |\psi\rangle \cdot \quad (34)
 \end{aligned}$$

It can be checked that the states $(1/2) \sum_u (-1)^{\langle v,u \rangle} |eu\rangle_{\text{env},C}$ are orthonormal for different e and v . As a result, the states $\sum_{eu} (1/2) \alpha_{ev} (-1)^{\langle v,u \rangle} |eu\rangle_{\text{env},C}$ are orthogonal for different v and have probability (square norm) given by $p_v = \sum_e |\alpha_{ev}|^2$. Introducing $|\tilde{v}\rangle_{\text{env},C} = \sum_{eu} (1/2) \alpha_{ev} (-1)^{\langle v,u \rangle} |eu\rangle_{\text{env},C}$, we can write the sum of Equation (34) as

$$\sum_v \left(\sum_{eu} \frac{1}{2} \alpha_{ev} (-1)^{\langle v,u \rangle} |eu\rangle_{\text{env},C} \right) \sigma_v |\psi\rangle = \sum_v \sqrt{p_v} |\tilde{v}\rangle_{\text{env},C} \sigma_v |\psi\rangle \cdot \quad (35)$$

showing that the twirled error model behaves like randomly applied Pauli matrices with σ_v applied with probability p_v . It is a recommended exercise to reproduce the above argument using the operator sum formalism.

To obtain the standard depolarizing error model with equal probabilities for the Pauli matrices, it is necessary to strengthen the randomization procedure by applying a random member U of the group generated by the 90° rotations around the x -, y -, and z -axis before the error and then undoing U by applying U^{-1} .

Randomization can be used to transform any one-qubit error model into the depolarizing error model. This explains why the depolarizing model is so useful for analyzing error correction techniques in situations in which errors act independently on different qubits. However, in many physical situations, the independence assumptions are not satisfied. For example, errors from common internal couplings between qubits are generally pairwise correlated to first order. In addition, the operations required to manipulate the qubits and to control the encoded information act on pairs at a time, which tends to spread even single-qubit errors. Still, in all these cases, the primary error processes are local. This means that there usually exists an environment-labeled sum expression for the total error process in which the amplitudes associated with errors acting simultaneously at k locations in time and space decrease exponentially with k . In such cases, error correction methods that handle all or most errors involving sufficiently few qubits are still applicable.

Quantum Error Analysis. One of the most important consequences of the subsystems interpretation of encoding quantum information in a physical system is that the encoded quantum information can be error-free even though errors have severely changed the state of the physical system. Almost trivially, any error operator acting only on the syndrome subsystem has no effect on the quantum information. The goal of error correction is to actively intervene and maintain the syndrome subsystem in states where the dominant error operators continue to have little effect on the information of interest. An important issue in analyzing error correction methods is to estimate the residual error in the encoded information. A simple example of how that can be done was discussed for the quantum repetition code. The same ideas can be applied in general. Let sys be the physical system in which the information is encoded, and $|\psi\rangle_{\text{sys}}$ an initial state containing such information with the syndrome subsystem appropriately prepared. Errors and error-correcting operations modify the state. The new state can be expressed with environment labeling as $\sum_e |e\rangle_{\text{env}} A_e^{(\text{sys})} |\psi\rangle_{\text{sys}}$. In view of the partitioning into information-carrying and syndrome subsystems, good states $|e\rangle_{\text{env}}$ are those states for which $A_e^{(\text{sys})}$ acts only on the syndrome subsystem, given that the syndrome has been prepared. The remaining states $|e\rangle$ form the set of bad states, B . The error probability p_e can be bounded from above by

$$\begin{aligned}
 p_e &\leq \left| \sum_{e \in B} |e\rangle_{\text{env}} A_e^{(\text{sys})} |\psi\rangle_{\text{sys}} \right|^2 \\
 &\leq \left(\sum_{e \in B} |e\rangle_{\text{env}} |A_e^{(\text{sys})}|_1 \right)^2,
 \end{aligned} \tag{36}$$

where $|A|_1 = \max_\phi \langle \phi | A | \phi \rangle$, the maximum being taken over normalized states. The second inequality usually leads to a gross overestimate but is independent of the encoded information and often suffices for obtaining good results. Because the environment-labeled

sum is not unique, a goal of the representation of the errors acting on the system is to use “good” operators to the largest extent possible. The flexibility of these error expansions makes them very useful for analyzing error models in conjunction with error correction methods.

In principle, we can obtain better expressions for p_e by calculating the density matrix ρ of the state of the subsystem containing the desired quantum information. This calculation involves tracing over the syndrome subsystem. The matrix ρ can then be compared to the intended state. If the intended state is pure, given by $|\phi\rangle$, the probability of error is given by $1 - \langle\phi|\rho|\phi\rangle$, which is the probability that a measurement that distinguishes between $|\phi\rangle$ and its orthogonal complement fails to detect $|\phi\rangle$. The quantity $\langle\phi|\rho|\phi\rangle$ is called the fidelity of the state ρ .

For applications to communication, the goal is to be able to reliably transmit arbitrary states through a communication channel, which may be physical or realized via an encoding/decoding scheme. It is therefore important to characterize the reliability of the channel independent of the information transmitted. Equation (36) can be used to obtain state-independent bounds on the error probability but does not readily provide a single measure of reliability. One way to quantify the reliability is to identify the error of the channel with the average error ϵ_a over all possible input states. The reliability is then given by the average fidelity $1 - \epsilon_a$. Another elegant way appropriate for QIP is to use the entanglement fidelity (Schumacher 1996). Entanglement fidelity measures the error when the input is maximally entangled with an identical reference system. In this process, the reference system is imagined to be untouched, so that the state of the reference system, together with the output state, can be compared with the original entangled state. For a one-qubit channel labeled sys, the reference system is a qubit, which we label “ref.” An initial, maximally entangled state is

$$|B\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle_{\text{ref}} |0\rangle_{\text{sys}} + |1\rangle_{\text{ref}} |1\rangle_{\text{sys}} \right) . \tag{37}$$

The reference qubit is assumed to be perfectly isolated and not affected by any errors. The final state $\rho^{(\text{ref},\text{sys})}$ is compared with $|B\rangle$, which gives the entanglement fidelity according to the formula $f_e = \langle B|\rho^{(\text{ref},\text{sys})}|B\rangle$. The entanglement error is $\epsilon_e = 1 - f_e$. It turns out that this definition does not depend on the choice of maximally entangled state. Fortunately, the entanglement error and the average error ϵ_a are related by a linear expression:

$$\epsilon_a = \frac{2}{3} \epsilon_e . \tag{38}$$

For k -qubit channels, the constant $2/3$ is replaced by $2^k/(2^k + 1)$. Experimental measurements of these fidelities do not require the reference system. There are simple averaging formulas to express them in terms of the fidelities for transmitting each of a sufficiently large set of pure states. An example of the experimental determination of the entanglement fidelity when the channel is realized by error correction is provided in Knill et al. (2001).

From Quantum Error Detection to Error Correction

In the independent depolarizing error model with small probability p of depolarization, the most likely errors are those that affect a small number of qubits. That is, if we define the weight of a product of Pauli operators to be the number of qubits affected, the dominant errors are those of small weight. Because the probability of a nonidentity Pauli operator is $3p/4$ —see Equation (31)—one expects about $(3p/4)n$ of n qubits to be changed. As a result, good error-correcting codes are considered to be those for which all errors of weight $\leq e \equiv (3p/4)n$ can be corrected. It is desirable that e have a high rate, which means that it is a large fraction of the total number of qubits n (the length of the code). Combinatorially, good codes are characterized by a high minimum distance, a concept that arises naturally in the context of error detection.

Quantum Error Detection. Let C be a quantum code, that is, a subspace of the state space of a quantum system. Let P be the operator that projects onto C , and $P^\perp = \mathbb{1} - P$ the one that projects onto the orthogonal complement. Then the pair P, P^\perp is associated with a measurement that can be used to determine whether a state is in the code or not. If the given state is $|\psi\rangle$, the result of the measurement is $P|\psi\rangle$ with probability $|P|\psi\rangle|^2$ and $P^\perp|\psi\rangle$ otherwise. As in the classical case, an error-detection scheme consists of preparing the desired state $|\psi_i\rangle \in C$, transmitting it through, say, a quantum channel, then measuring whether the state is still in the code, accepting the state if it is, and rejecting it otherwise. We say that C detects error operator E if states accepted after E had acted are unchanged except for an overall scale. Using the projection operators, this is the statement that for every state $|\psi_i\rangle \in C$, $PE|\psi_i\rangle = \lambda_E |\psi_i\rangle$. Because $P|\psi\rangle$ is in the code for every $|\psi\rangle$, it follows that $PEP|\psi\rangle = \lambda_E P|\psi\rangle$. It follows that a characterization of detectability is given by Theorem 3.

Theorem 3. E is detectable by C if and only if $PEP = \lambda_E P$ for some λ_E .

A second characterization is given by Theorem 4.

Theorem 4. E is detectable by C if and only if for all $|\psi\rangle, |\phi\rangle \in C$, $\langle \psi | E | \phi \rangle = \lambda_E \langle \psi | \phi \rangle$ for some λ_E .

A third characterization is obtained by taking the condition for classical detectability in Theorem 1 and replacing \neq by orthogonal to:

Theorem 5. E is detectable by C if and only if for all $|\phi\rangle, |\psi\rangle$ in the code with $|\phi\rangle$ orthogonal to $|\psi\rangle$, $E|\phi\rangle$ is orthogonal to $|\psi\rangle$.

For a given code C , the set of detectable errors is closed under linear combinations. That is, if E_1 and E_2 are both detectable, then so is $\alpha E_1 + \alpha E_2$. This useful property implies that, to check detectability, one has to consider only the elements of a linear basis for the space of errors of interest.

Consider n -qubits with independent depolarizing errors. A robust error-detecting code should detect as many of the small-weight errors as possible. This requirement motivates the definition of minimum distance: The code C has minimum distance d if the smallest-weight product of Pauli operators E for which C does not detect E is d . The notion comes from classical codes for bits, where a set of code words C' has minimum distance d if the

smallest number of flips required to change one code word in C' into another one in C' is d . For example, the repetition code for three bits has minimum distance 3. Note that the minimum distance for the quantum repetition code is 1: Applying $\sigma_z^{(1)}$ preserves the code and changes the sign of $|111\rangle$ but not of $|000\rangle$. As a result, $\sigma_z^{(1)}$ is not detectable. The notion of minimum distance can be generalized for error models with specified first-order error operators (Knill et al. 2000). In the case of depolarizing errors, the first-order error operators are single-qubit Pauli matrices, which are the errors of weight 1.

Quantum Error Correction. Let $\mathcal{E} = \{E_0 = \mathbb{1}, E_1, \dots\}$ be the set of errors that we wish to be able to correct. When a decoding procedure for the code C exists such that all errors in \mathcal{E} are corrected, we say that \mathcal{E} is correctable (by C). A situation in which correctability of \mathcal{E} is apparent occurs when the errors E_i are unitary operators satisfying the condition that $E_i C$ are mutually orthogonal subspaces. The repetition code has this property for the set of errors consisting of the identity and Pauli operators acting on a single qubit. In this situation, the procedure for decoding is to first make a projective measurement and determine which of the subspaces $E_i C$ the state is in and then to apply the inverse of the error operator, that is, E_i^\dagger . This situation is not far from the generic one. One characterization of correctability is described in Theorem 6.

Theorem 6. \mathcal{E} is correctable if and only if there is a linear transformation of the set \mathcal{E} such that the operators E'_i in the new set satisfy the following properties: (1) The $E'_i C$ are mutually orthogonal, and (2) E'_i restricted to C is proportional to a restriction to C of a unitary operator.

To relate this characterization to detectability, note that the two properties imply that $(E'_i)^\dagger E'_j C$ is orthogonal to C if $i \neq j$ and $(E'_i)^\dagger E'_i$ restricted to C is proportional to the identity on C . In other words, the $(E'_i)^\dagger E'_i$ are detectable. This detectability condition applied to the original error set constitutes a second characterization of correctability, as given in Theorem 7.

Theorem 7. \mathcal{E} is correctable if and only if the operators in the set $\mathcal{E}^\dagger \mathcal{E} = \{E_1^\dagger E_2 : E_i \in \mathcal{E}\}$ are detectable.

Before explaining the characterizations of correctability, we consider the situation of n qubits, where the characterization by detectability (Theorem 7) leads to a useful relationship between minimum distance and correctability of low-weight errors.

Theorem 8. If a code on n qubits has a minimum distance of at least $2e + 1$, then the set of errors of weight at most e is correctable.

This theorem follows by observing that the weight of $E_1^\dagger E_2$ is at most the sum of the weights of the E_i . As a result of this observation, the problem of finding good ways to correct all errors up to a maximum weight reduces to that of constructing codes with sufficiently high minimum distance. Thus, questions such as “what is the maximum dimension of a code of minimum distance d on n qubits?” are of great interest. As in the case of classical coding theory, this problem appears to be very difficult in general. Answers are known for small n (Calderbank et al. 1998), and there are asymptotic bounds (Ashikhmin and Litsyn 1999). Of course, for achieving low error probabilities, it is not necessary to correct all errors of weight $\leq e$, just almost all such errors. For example, the concatenated codes used for fault-tolerant quantum computation achieve this goal (see “Fault-Tolerant Quantum Communication and Computation” later in this article).

For the remainder of this section, we explain the characterizations of correctability. Using the conditions for detectability from the previous section, the condition for correctability in Theorem 7 is equivalent to

$$PE_i^\dagger E_j P = \lambda_{ij} P . \quad (39)$$

This condition is preserved under a linear change of basis for \mathcal{E} . That is, if A is any invertible matrix with coefficients a_{ij} , we can define new error operators $D_k = \sum_i E_i a_{ik}$. For the D_k , the left side of Equation (39) is

$$\begin{aligned} PD_k^\dagger D_l P &= P \left(\sum_{ij} \bar{a}_{ik} E_i^\dagger E_j a_{jl} \right) P \\ &= \sum_{ij} \bar{a}_{ik} a_{jl} \lambda_{ij} P \\ &= (A^\dagger \Lambda A)_{kl} P , \end{aligned} \quad (40)$$

where Λ is the matrix formed from the λ_{ij} . Using the fact that Λ is a positive semidefinite matrix (that is, for all x , $x^\dagger \Lambda x \geq 0$, and $\Lambda^\dagger = \Lambda$), we can choose A such that $A^\dagger \Lambda A$ is

of the form $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. In this matrix, the upper left block is the identity operator for

some dimension.

An important consequence of invariance under a change of basis of error operators is that the set of errors correctable by a particular code and decoding procedure is linearly closed. Thus, if E and D are corrected by the decoding procedure, then so is $\alpha E + \beta D$. This observation also follows from the linearity of quantum mechanically implementable operations.

We explain the condition for correctability by using the subsystems interpretation of decoding procedures. For simplicity, assume that $1 \in \mathcal{E}$. To show that correctability of \mathcal{E} implies detectability of all $E \in \mathcal{E}^\dagger \mathcal{E}$, suppose that we have a decoding procedure that recovers the information encoded in C after any of the errors in the set \mathcal{E} have occurred. Every physically realizable decoding procedure can be implemented by first adding ancilla quantum systems in a prepared pure state to form a total system labeled T , then applying a unitary map U to the state of T , and finally separating T into a pair of systems (syn, Q), where “syn” corresponds to the syndrome subsystem and Q is a quantum system with the same dimension as the code that carries the quantum information after decoding. Denote the state space of the physical system containing C as \mathcal{H} and the state space of system X by \mathcal{H}_X , where X is any one of the other systems. Let V be the unitary operator that encodes information by mapping \mathcal{H}_Q onto $C \subseteq \mathcal{H}$. We have the following relationships:

$$\mathcal{H}_Q \xleftrightarrow{V} C \subseteq \mathcal{H} \subseteq \mathcal{H}_T \xleftrightarrow{U} \mathcal{H}_{\text{syn}} \otimes \mathcal{H}_Q . \quad (41)$$

Here, we used bidirectional arrows to emphasize that the operators V and U can be inverted on their range and therefore identify the states in their domains with the states in their ranges. The inclusion $\mathcal{H} \subseteq \mathcal{H}_T$ implicitly identifies \mathcal{H} with the subspace determined by the prepared pure state on the ancillas. The last state space of Equation (41) is expressed as a tensor product, which is the state space of the combined system (syn, Q).

For states of \mathcal{H}_Q , we will write $|\psi\rangle = |\psi\rangle_Q \leftrightarrow |\psi\rangle_L \in C$. Because $\mathbb{1}$ is a correctable error, it must be the case that $|\psi\rangle_L \xleftrightarrow{U} |0\rangle_{\text{syn}}|\psi\rangle \in \mathcal{H}_{\text{syn}} \otimes \mathcal{H}_Q$ for some state $|0\rangle_{\text{syn}}$. To establish this fact, use the linearity of the maps. In general,

$$\begin{aligned} |\psi\rangle_L &\rightarrow E_i|\psi\rangle_L \\ &\xleftrightarrow{U} |i\rangle_{\text{syn}}|\psi\rangle . \end{aligned} \tag{42}$$

The $|i\rangle_{\text{syn}}$ need not be normalized or orthogonal. Let F be the subspace spanned by the $|i\rangle_{\text{syn}}$. Then U induces an identification of $F \otimes \mathcal{H}_Q$ with a subspace $\bar{C} \subseteq \mathcal{H}$. This is the desired subsystem identification. We can then see how the errors act in this identification.

$$\begin{aligned} |\psi\rangle_L &\leftrightarrow |0\rangle_{\text{syn}}|\psi\rangle \\ \downarrow & \\ E_i|\psi\rangle_L &\leftrightarrow |i\rangle_{\text{syn}}|\psi\rangle . \end{aligned} \tag{43}$$

This means that for all $|\psi\rangle$ and $|\phi\rangle$,

$${}^L\langle\psi|E_j^\dagger E_i|\phi\rangle_L = {}^{\text{syn}}\langle j|i\rangle_{\text{syn}}\langle\psi|\phi\rangle , \tag{44}$$

that is, all errors in $\mathcal{E}^\dagger\mathcal{E}$ are detectable.

Now, suppose that all errors in $\mathcal{E}^\dagger\mathcal{E}$ are detectable. To see that correctability of \mathcal{E} follows, choose a basis for the errors so that $\lambda_{ij} = \delta_{ij}\lambda_i$ with $\lambda_i = 1$ for $i < s$ and $\lambda_i = 0$ otherwise. Define a subsystem identification by

$$|i\rangle_{\text{sys}}|\psi\rangle \xrightarrow{W} E_i|\psi\rangle_L , \tag{45}$$

for $0 \leq i < s$. By assumption and construction, ${}^L\langle\psi|E_j^\dagger E_i|\psi\rangle_L = \delta_{ij}$, which implies that W is unitary (after linear extension), and so this is a proper identification. For $i \geq s$, $E_i|\psi\rangle_L = 0$, which implies that for states in the code, these errors have probability 0. Therefore, the identification can be used to successfully correct \mathcal{E} .

Constructing Codes

Stabilizer Codes. Most useful quantum codes are based on stabilizer constructions (Gottesman 1996, Calderbank et al. 1997). Stabilizer codes are useful because they make it easy to determine which Pauli-product errors are detectable and because they can be interpreted as special types of classical, linear codes. The latter feature makes it possible to use well-established techniques from the theory of classical error-correcting codes to construct good quantum codes.

A stabilizer code of length n for k -qubits (abbreviated as an $[[n, k]]$ code), is a 2^k -dimensional subspace of the state space of n -qubits that is characterized by the set of

products of Pauli operators that leave each state in the code invariant. Such Pauli operators are said to stabilize the code. A simple example of a stabilizer code is the quantum repetition code introduced earlier. The code's states $\alpha|000\rangle + \beta|111\rangle$ are exactly the states that are unchanged after applying $\sigma_z^{(1)} \sigma_z^{(2)}$ or $\sigma_z^{(1)} \sigma_z^{(3)}$. To simplify the notation, we write $I = 1$, $X = \sigma_x$, $Y = \sigma_y$, and $Z = \sigma_z$. A product of Pauli operators can then be written as $ZIXI = \sigma_z^{(1)} \sigma_x^{(3)}$ (as an example of length 4) with the ordering determining which qubit is being acted upon by the operators in the product.

We can understand the properties of stabilizer codes by working out the example of the quantum repetition code with the stabilizer formalism. A stabilizer of the code is $S = \{ZZI, ZIZ\}$. Let \bar{S} be the set of Pauli products that are expressible up to a phase as products of elements of S . For the repetition code, $\bar{S} = \{III, ZZI, ZIZ, IZZ\}$. \bar{S} consists of all Pauli products that stabilize the code. The crucial property of S is that its operators commute, that is, for $A, B \in S$, $AB = BA$. According to results from linear algebra, it follows that the state space \mathcal{H} can be decomposed into orthogonal subspaces \mathcal{H}_λ such that for $A \in S$ and $|\psi\rangle \in \mathcal{H}_\lambda$, $A|\psi\rangle = \lambda(A)|\psi\rangle$. The \mathcal{H}_λ are the common eigenspaces of S . The stabilizer code C defined by S is the subspace stabilized by the operators in S , which means that it is given by \mathcal{H}_λ with $\lambda(A) = 1$. The subspaces for other $\lambda(A)$ have equivalent properties and are often included in the set of stabilizer codes. For the repetition code, the stabilized subspace is spanned by the logical basis $|000\rangle$ and $|111\rangle$. From the point of view of stabilizers, there are two ways in which a Pauli product B can be detectable: (1) if $B \in \bar{S}$ because, in this case, B acts as the identity on the code and (2) if B anticommutes with at least one member (say A) of S . To see that this statement is correct, let $|\psi\rangle$ be in the code. Then $A(B|\psi\rangle) = (AB)|\psi\rangle = -(BA)|\psi\rangle = -B(A|\psi\rangle) = -B|\psi\rangle$. Thus, $B|\psi\rangle$ belongs to \mathcal{H}_λ with $\lambda(A) = -1$. Because this subspace is orthogonal to $C = \mathcal{H}_1$, B is detectable. We define the set of Pauli products that commute with all members of S as \bar{S}^\perp . Thus, B is detectable if either $B \notin \bar{S}^\perp$ or $B \in \bar{S}$. Note that because \bar{S} consists of commuting operators, $\bar{S} \subseteq \bar{S}^\perp$.

To construct a stabilizer code that can correct all errors of weight at most one (a quantum one-error-correcting code), it suffices to find S with the minimum weight of nonidentity members of \bar{S}^\perp being at least three ($3 = 2 \cdot 1 + 1$)—also refer to Theorem 8. In this case, we say that \bar{S}^\perp has minimum distance 3. As an example, we can exhibit a stabilizer for the famous length-five one-error-correcting code for one qubit (Bennett et al. 1996, Laflamme et al. 1996):

$$S = \{XZZXI, IXZZX, XIXZZ, ZXIXZ\} . \tag{46}$$

As a general rule, it is desirable to exhibit the stabilizer minimally, which means that no member is the product up to a phase of some of the other members. In this case, the number of qubits encoded is $n - |S|$, where n is the length of the code and $|S|$ is the number of elements of S .

To obtain the correspondence between stabilizer codes and classical binary codes, we replace the symbols I, X, Y , and Z in a Pauli product by 00, 01, 10, and 11, respectively. Thus, the members of the stabilizer can be thought of as binary vectors of length $2n$. We use arithmetic modulo 2 for sums, inner products, and application of a binary matrix. Because the numbers modulo 2 (\mathbb{Z}_2) form a mathematical field, the basic properties of vector spaces and linear algebra apply to binary vectors and matrices. Thus, the stabilizer is minimal in the sense introduced above if the corresponding binary vectors are independent over \mathbb{Z}_2 . Given two binary (column) vectors x and y of length 2 associated with Pauli products, the property of anticommuting is equivalent to $x^T B y = 1$, where B is the block diagonal $2n \times 2n$ matrix with 2×2 blocks given by $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

This means that \bar{S}^\perp can be identified with the set of vectors x such that $x^T B y = 0$ for all binary vectors y associated with the members of S . It turns out that the inner product $\langle x, y \rangle = x^T B y$ arises in the study of classical codes over the four-element mathematical field $GF(4)$, which can be represented by the vectors 00, 01, 10, and 11 with addition modulo 2 and a new multiplication operation. This relationship leads to the construction of many good stabilizer codes (Calderbank et al. 1998).

Conserved Quantities, Symmetries, and Noiseless Subsystems. Even though a physical system may be exposed to error, some of its properties are often not affected by the errors. If these conserved quantities can be identified with the defining quantities of qubits or other information units, error-free storage of information can be ensured without active intervention. This is the idea behind noiseless subsystems.

When do noiseless subsystems exist and how can they be constructed? The examples discussed in the previous sections show that a noiseless subsystem may be a subset of physical qubits, as in the trivial two-qubit example, or it may require a more abstract subsystem identification, as in the example of the three spin-1/2 particles. As will be explained, in both cases, there are quantities conserved by the errors that can be used to identify the noiseless subsystem.

A simple classical example for the use of conserved quantities consists of two physical bits subject to errors that either flip both bits or leave them alone. A quantity invariant under this noise model is the parity $P(s)$ of a state s of the two bits. The parity $P(s)$ is defined as the number of 1s in the bit string s reduced modulo 2: $P(00) = P(11) = 0$, and $P(01) = P(10) = 1$. Flipping both bits does not change the value of P . Consequently, the two values of P can be used to identify the two states of a noiseless bit. The syndrome subsystem can be associated with the value (nonconserved) of the first physical bit using the function defined by $F(0b) = 0$, $F(1b) = 1$. The corresponding subsystem identification is obtained by using the values of P and F as the states of the syndrome (left) and the noiseless information-carrying subsystem (right) according to $ab \leftrightarrow F(ab) \cdot P(ab)$.

In quantum systems, conserved quantities are associated with the presence of symmetries, that is, with operators that commute with all possible errors. In the trivial two-qubit example, operators acting only on qubit 2 commute with the error operators. In particular, if E is any one of the errors, $E\sigma_u^{(2)} = \sigma_u^{(2)}E$ for $u = x, y, z$. It follows that the expectations of $\sigma_u^{(2)}$ are conserved. That is, if ρ is the initial state (density matrix) of the two physical qubits and ρ' is the state after the errors acted, then $\text{tr} \sigma_u^{(2)} \rho' = \text{tr} \sigma_u^{(2)} \rho$. Because the state of qubit 2 is completely characterized by these expectations, it follows immediately that it is unaffected by the noise.

The trivial two-qubit example suggests a general strategy for finding a noiseless qubit: First, determine the commutant of the errors, which is the set of operators that commute with all errors. Then, find a subset of the commutant that is algebraically equivalent to the operators characterizing a qubit. The equivalence can be formulated as a one-to-one map f from qubit operators to operators in the commutant. For the range of f to be algebraically equivalent, f must be linear and satisfy $f(A^\dagger) = f(A)^\dagger$ and $f(AB) = f(A)f(B)$. Once such an equivalence is found, a fundamental theorem from the representation theory of finite dimensional operator algebras implies that a subsystem identification for a noiseless qubit exists (Knill et al. 2000, Viola et al. 2001).

The strategy can be applied to the example of three spin-1/2 particles subject to collective errors. One can determine the commutant by using the physical properties of spin to find the conserved quantities associated with operators in the commutant, as suggested in the box “Creating a Noiseless Subsystem from Three Spin-1/2 Particles” on page 205. Alternatively, observe that, by definition, this error model is symmetric under permutations of the particles. Therefore, the actions of these permutations on the state

space form a group Π of unitary operators commuting with the errors. It is a fact that the commutant of the set of collective errors consists of the linear combinations of operators in Π . With respect to the group Π , one can immediately determine the space $V_{3/2}$ of symmetric states, that is, those that are invariant under the permutations. It is spanned by

$$|\uparrow\uparrow\uparrow\rangle, \frac{1}{\sqrt{3}}(|\uparrow\uparrow\downarrow\rangle + |\uparrow\downarrow\uparrow\rangle + |\downarrow\uparrow\uparrow\rangle), \frac{1}{\sqrt{3}}(|\uparrow\downarrow\downarrow\rangle + |\downarrow\uparrow\downarrow\rangle + |\downarrow\downarrow\uparrow\rangle), |\downarrow\downarrow\downarrow\rangle. \quad (47)$$

A basic result from the representation theory of groups implies that the projection onto $V_{3/2}$ is given by $P_{3/2} = (1/6)\sum_{g \in \Pi} g$. The orthogonal complement $V_{1/2}$ of $V_{3/2}$ is invariant under Π and can be analyzed separately. With the subsystem identification of Equation (26) already in hand, one can see that the permutation π_1 , which permutes the spins according to $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$, acts on the noiseless qubit, by applying $Z_{240^\circ} = e^{-i\sigma_z 2\pi/3}$, a 240° rotation around the z -axis. Similarly, the permutation π_2 , which exchanges the last two spins, acts as σ_x on the qubit. To make them algebraically equivalent to the corresponding qubit operators, it is necessary to eliminate their action on $V_{3/2}$ by projecting onto $V_{1/2}$: $\pi'_1 = (1 - P_{3/2})\pi_1$ and $\pi'_2 = (1 - P_{3/2})\pi_2$. Sums of products of π'_1 and π'_2 are equivalent to the corresponding sums of products of Z_{240° and σ_x , which generate all qubit operators. To get the subsystem identification of Equation (26), one can start with a common eigenstate $|\psi\rangle$ of π'_1 (a z -rotation on the noiseless qubit) and $2J_z$ (the syndrome subsystem's σ_z) with eigenvalues $e^{-i2\pi/3}$ and 1, respectively. The choice of eigenvalues implies that $|\psi\rangle \leftrightarrow |\uparrow\rangle \cdot |\circ\rangle$ in the desired identification. We can obtain the other logical states of the syndrome spin 1/2 and the noiseless qubit by applying π'_2 , $2J_x$, and $\pi'_2 2J_x$ to $|\psi\rangle$, which act by flipping the states of the qubit or the syndrome spin. This method for obtaining the subsystem identification generalizes to other operator equivalences and error operators.

Fault-Tolerant Quantum Communication and Computation

The utility of information and information processing depends on the ability to implement large numbers of information units and information-processing operations. We say that an implementation of information processing is scalable if the implementation can realize arbitrarily many information units and operations without loss of accuracy and with physical resource overheads that are polynomial (or efficient) in the number of information units and operations. Scalable information processing is achieved by implementing information fault-tolerantly.

One of the most important results of the work in quantum error-correction and fault-tolerant computation is the accuracy threshold theorem, according to which scalability is possible, in principle, for quantum information.

Theorem 9. Assume the requirements for scalable QIP (see below). If the error per gate is less than a threshold, then it is possible to efficiently quantum-compute to arbitrary accuracy.

Requirements for Scalable QIP. The value of the threshold accuracy (or error) depends strongly on which set of requirements is used—in particular, the error model that is assumed. The requirements are closely related to the basic requirements for constructing a quantum information processor (DiVincenzo 2000) but have to include

explicit assumptions on the error model and on the temporal and spatial aspects of the available quantum control:

Scalable physical systems. It is necessary to access physical systems that are able to support qubits or other basic units of quantum information. The systems must be scalable; that is, they must be able to support any number of independent qubits.

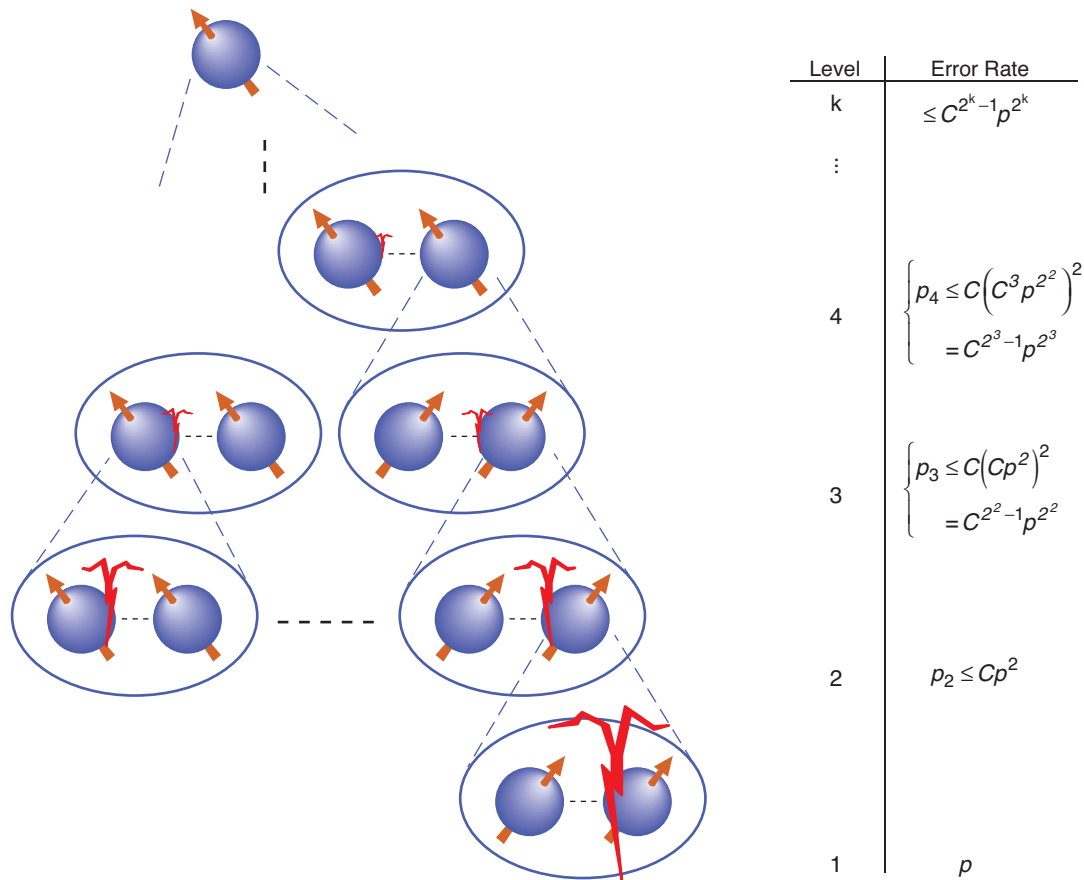
State preparation. One must be able to prepare any qubit in the standard initial state $|0\rangle$. Any preexisting content is assumed to be lost, as would happen if, for example, the qubit is first discarded and then replaced by a prepared one. The condition can be weakened; that is, it is sufficient that a large fraction of the qubits can be prepared in this way.

Measurement. Being able to measure any qubit in the logical basis is a requirement. Again, it is sufficient that a large enough fraction of the qubits are measurable. For solving computational problems with deterministic answers, the standard projective measurement can be replaced by weak measurements that return a noisy number whose expectation is the probability that a qubit is in the state $|1\rangle$ (Laflamme et al. 2001).

Quantum control. One must be able to implement a universal set of unitary quantum gates acting on a small number (usually, at most, two at a time) of qubits. For most accuracy thresholds, it is necessary to be able to apply the quantum control in parallel to any number of disjoint pairs of qubits. This parallelism requirement can be weakened if a nearly noiseless quantum memory is available. The requirement that it be possible to apply two-qubit gates to any pair of qubits is unrealistic given the constraints of three-dimensional space. Work on how to deal with this problem is ongoing (Aharonov and Ben-Or 1999). The universality assumption can be substantially weakened by replacement of some or all unitary quantum gates with operations to prepare special states or by additional measurement capabilities. See, for example, Michael Nielsen (2001) and the references therein.

Errors. The error probability per gate must be below a threshold and satisfy independence and locality properties (refer to the section “Error Models”). The definition of gate includes the “no-op,” which is the identity operation implemented over the time required for a computational step. For the most pessimistic, independent, local error models, the error threshold is above $\sim 10^{-6}$; for the independent depolarizing errors, it is believed to be better than 10^{-4} (Gottesman and Preskill 1999). For some special error models, the threshold is substantially higher. For example, for the independent “erasure” error model, where error events are always detected, the threshold is above .01, and for an error model whose errors are specific, unintentional measurements in the standard basis of a qubit, the threshold is 1 (Knill et al. 2000). The threshold is also well above .01 when the goal is only to transmit quantum information through noisy quantum channels (Briegel et al. 1998).

Realizing Fault Tolerance. The existing proofs of the accuracy threshold theorems consist of explicit instructions for building a scalable quantum information processor and analyses of its robustness against the assumed error model. The instructions for realizing scalable computation are based on the following simple idea. Suppose that the error rate per operation for some way of realizing qubits is p . We can use these qubits and a quantum error-correcting code to encode logical qubits for which the storage error rate is reduced. For example, if a one-error correcting code is used, the error rate per storage interval for the logical qubits is expected to be $\leq cp^2$ for some constant c . Suppose that we can show how to implement encoded operations, preparations, measurement, and the subroutines required for error correction such that this inequality is now valid for each basic encoded step, perhaps for a larger constant C . Suppose furthermore that the errors for the encoded information still satisfy the assumed error model. The newly defined logical qubits then have an error rate of $\leq Cp^2$, which is less than p



for $p < 1/C$. We can use the newly realized qubits as a foundation for making higher-level logical qubits. The result is multiple levels of encodings. In the next level (level 2), the error rate is $\leq C^3p^4$, and after k iterations, it is $\leq C^{2^k-1}p^{2^k}$, a doubly exponentially decreasing function of k . This procedure is called concatenation (refer to Figure 8). Because the complexity, particularly the number of physical qubits needed for each final logical qubit, grows only singly exponentially in k , the procedure is efficient. Specifically, to achieve a logical error of ϵ per operation requires of the order of $|\log(\epsilon)|^r$ resources per logical qubit for some finite r . In practice, this simple idea is still dauntingly complex, but there is hope that, for realistic errors in physical systems and by cleverly trading off different variations of these techniques, much of the theoretical complexity can be avoided (Steane 1999).

Many important developments and ideas of quantum information were ultimately needed to realize encoded operations, preparations, measurements, and error correction subroutines that behave well with respect to concatenation. Stabilizer codes provide a particularly nice setting for implementing many of these techniques. One reason is that good stabilizer codes are readily constructed. Another is that they enable encoding operations in a way that avoids spreading errors between the qubits of a single code word (Gottesman 1998). In addition, there are many tricks based on teleportation that can be used to maintain the syndrome subsystems in acceptably low error states and to implement general operations systematically (Gottesman and Chuang 1999). To learn more about all these techniques, see the textbook by Nielsen and Isaac Chuang (2001) and the works of Daniel Gottesman (1998) and John Preskill (1998).

Figure 8. Schematic Representation of Concatenation

The bottom level represents qubits realized more or less directly in a physical system. Each next level represents logical qubits defined by means of subsystems in terms of the previous level's qubits. More efficient subsystems might represent multiple qubits in one code block rather than the one qubit per code block shown here.

Concluding Remarks

The advancements in quantum error-correction and fault-tolerant QIP have shown that, in principle, scalable quantum computation is achievable. This is a crucial result because it suggests that experimental efforts in QIP will eventually lead to more than a few small-scale applications of quantum information to communication and problems with few qubits. However, the general techniques for achieving scalability that are known are difficult to realize. Existing technologies are far from achieving sufficient accuracy even for just two qubits—at least in terms of the demands of the usual accuracy-threshold theorems. There is hope that more optimistic thresholds can be shown to apply if one takes into consideration the specific constraints of a physical device, better understands the dominant sources of errors, and exploits tailor-made ways of embedding quantum information into subsystems. Current work in this area is focused on finding such methods of quantum error control. These methods include approaches to error control not covered in this article—for example, techniques for actively turning off the error-inducing environmental interactions (Viola and Lloyd 1998, Viola et al. 1999) and modifications to controlling quantum systems that eliminate systematic and calibration errors (Levitt 1982, Cummins and Jones 1999). Further work is also needed to improve the thresholds for the more pessimistic error models and for developing more-efficient scalability schemes. ■

Contact Information

E. Knill: knill@lanl.gov

R. Laflamme: laflamme@iqc.ca

A. Ashikhmin:
aea@research.bell-labs.com

H. Barnum: barnum@lanl.gov

L. Viola: viola@lanl.gov

W. H. Zurek: whz@lanl.gov

Further Reading

- Aharonov, D., and M. Ben-Or. 1996. Fault-Tolerant Quantum Computation with Constant Error Rate. In *Proceedings of the 29th Annual ACM Symposium on the Theory of Computation (STOC)*. New York: ACM Press.
- . 1999. Fault-Tolerant Quantum Computation with Constant Error. [Online]: <http://eprints.lanl.gov/quant-ph/9906129>.
- Ashikhmin, A., and S. Litsyn. 1999. Upper Bounds on the Size of Quantum Codes. *IEEE Trans. Inf. Theory* **45**: 1206.
- Bennett, C. H., D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. 1996. Mixed State Entanglement and Quantum Error-Correcting Codes. *Phys. Rev. A* **54**: 3824.
- Braunstein, S. L., and H.-K. Lo, eds. 2000. Special Focus Issue on Experimental Proposals for Quantum Computation: Forward. *Fortschr. Phys.* **48** (9–11): 76
- Briegel, H.-J., W. Dür, J. I. Cirac, and P. Zoller. 1998. Quantum Repeaters for Communication. [Online]: <http://eprints.lanl.gov/quant-ph/9803056>.
- Calderbank, A. R., E. M. Rains, P. W. Shor, and N. J. A. Sloane. 1997. Quantum Error Correction and Orthogonal Geometry. *Phys. Rev. A* **78**: 405.
- . 1998. Quantum Error Correction via Codes over $gf(4)$. *IEEE Trans. Inf. Theory* **44**: 1369.
- Cummins, H. K., and J. A. Jones. 2000. Use of Composite Rotations to Correct Systematic Errors in NMR Quantum Computation. *New J. Phys.* **2**: 6.
- DiVincenzo, D. P. 2000. The Physical Implementation of Quantum Computation. *Fortschr. Phys.* **48**: 771.
- Gottesman, D. 1996. A Class of Quantum Error-Correcting Codes Saturating the Quantum Hamming Bound. *Phys. Rev. A* **54**: 1862.
- . 1998. A Theory of Fault-Tolerant Quantum Computation. *Phys. Rev. A* **57**: 127.
- Gottesman, D., and I. L. Chuang. 1999. Demonstrating the Viability of Universal Quantum Computation Using Teleportation and Single-Qubit Operations. *Nature* **402**: 390.
- Kitaev, A. Yu. 1997. Quantum Error Correction with Imperfect Gates. In *Quantum Communication and Computing and Measurement*. Edited by O. Hirota et al. New York: Plenum.
- Knill, E., and R. Laflamme. 1996. “Concatenated Quantum Codes.” [Online]: <http://eprints.lanl.gov/quant-ph/9608012>.

- Knill, E., R. Laflamme, and G. Milburn. 2000. "Thresholds for Linear Optics Quantum Computation." [Online]: [http://eprints.lanl.gov. \(quant-ph/0006120\)](http://eprints.lanl.gov. (quant-ph/0006120)).
- Knill, E., R. Laflamme, and W. H. Zurek. 1998a. Resilient Quantum Computation. *Science* **279**: 342.
- . 1998b. Resilient Quantum Computation: Error Models and Thresholds. *Proc. R. Soc. London, Ser. A* **454**: 365.
- Knill, E., R. Laflamme, and L. Viola. 2000. Theory of Quantum Error Correction for General Noise. *Phys. Rev. Lett.* **84**: 2525.
- Kraus, K. 1983. States, Effects and Operations: Fundamental Notions of Quantum Theory. *Lecture Notes in Physics*. Vol. 190. Berlin: Springer-Verlag.
- Laflamme, R., C. Miquel, J-P. Paz, and W. H. Zurek. 1996. Perfect Quantum Error-Correcting Code. *Phys. Rev. Lett.* **77**: 198.
- Levitt, M. H. 1982. Symmetrical Composite Pulse Sequences for NMR Population-Inversion. I. Compensation for Radiofrequency Field Inhomogeneity. *J. Mag. Res.* **48**:234.
- Nielsen, M. A. 2001. Universal Quantum Computation Using Only Projective Measurement, Quantum Memory, and Preparation of the $|0\rangle$ State. [Online]: [http://eprints.lanl.gov. \(quant-ph/0108020\)](http://eprints.lanl.gov. (quant-ph/0108020)).
- Nielsen, M. A., and I. L. Chuang. 2001. *Quantum Computation and Quantum Information*. Cambridge, U.K.: Cambridge University Press.
- Preskill, J. 1998. Reliable Quantum Computers. *Proc. R. Soc. London, Ser. A* **454**: 385.
- Schumacher, B. 1996. Sending Entanglement through Noisy Quantum Channels. *Phys. Rev. A* **54**: 2614.
- Shor, P. W. 1995. Scheme for Reducing Decoherence in Quantum Computer Memory. *Phys. Rev. A* **2**: 2493.
- . 1996. Fault-Tolerant Quantum Computation. In *Proceedings of the 37th Symposium on the Foundations of Computer Science (FOCS)*, p. 56. Los Alamitos, CA: IEEE Press.
- Steane, A., 1996. Multiple Particle Interference and Quantum Error Correction. *Proc. R. Soc. London, Ser. A* **452**: 2551.
- . 1999. Efficient Fault-Tolerant Quantum Computing. *Nature* **399**: 124.
- Viola, L., and S. Lloyd. Dynamical Suppression of Decoherence in Two-State Quantum Systems. *Phys. Rev. A* **58**: 2733.
- Viola, L., E. Knill, R. Laflamme. 2001. Constructing Qubits in Physical Systems. *J. Phys. A* **34** (35): 7067.
- Viola, L., E. Knill, and S. Lloyd. 1999. Dynamical Decoupling of Open Quantum Systems. *Phys. Rev. Lett.* **82**: 2417.

Glossary

- Bit.** The basic unit of deterministic information. It is a system that can be in one of two possible states, 0 and 1.
- Bit string.** A sequence of 0s and 1s that represents a state of a sequence of bits. The bit strings are words in the binary alphabet.
- Classical information.** The type of information based on bits and bit strings and, more generally, on words formed from finite alphabets. This is the information used for communication between people. Classical information can refer to deterministic or probabilistic information, depending on the context.
- Code.** A set of states that can be used to represent information. The set of states needs to have the properties of the type of information to be represented. The code is usually a subset of the states of a given system Q . It is then a Q -code or a code on Q . If information is represented by a state in the code, Q is said to carry the information.
- Code word.** A state in a code. The term is primarily used for classical codes defined on bits or systems with nonbinary alphabets.
- Concatenation.** An iterative procedure in which higher-level logical information units are implemented in terms of lower-level units.
- Control error.** An error due to nonideal control in applying operations or gates.
- Communication channel.** A means for transmitting information from one place to another. It can be associated with a physical system in which the information to be transmitted is stored by the sender. The system is subsequently conveyed to the receiver, who can then make use of the information.
- Correctable error set.** For a given code, a set of errors such that there is an implementable procedure R that, after any one of the errors E acts on a state x in the code, returns the system to the state $x = REx$. What procedures are implementable depends on the type of information represented by the system and, if it is a physical system, its physics.
- Decoding.** The process of transferring information from an encoded form to its “natural” form. In the context of error correction, decoding is often thought of as consisting of two steps: one which removes the errors’ effects (sometimes called the recovery procedure) and one that extracts the information (often also called decoding in a narrower sense).
- Depolarizing errors.** An error model for qubits in which random Pauli operators are applied independently to each qubit.
- Detectable error.** For a given code, an error that has no effect if the state is observed to have remained in the code. If the state is no longer in the code, the error is said to have been detected, and the state no longer represents valid information.
- Deterministic information.** The type of information based on bits and bit strings. This is the same as classical information but explicitly excludes probabilistic information.
- Encoding.** The process of transferring information from its natural form to an encoded form. It requires an identification of the valid states associated with the information and the states of a code. The process acts on an information unit and replaces it with the system whose state space contains the code.
- Environment.** In the context of information encoded in a physical system, it refers to other physical systems that may interact with the information-carrying system.
- Environmental noise.** Noise due to unwanted interactions with the environment.
- Error.** Any unintended effect on the state of a system, particularly in storing or otherwise processing information.

- Error basis.** A set of state transformations that can be used to represent any error. For quantum systems, errors can be represented as operators acting on the system's state space, and an error basis is a maximal, linearly independent set of such operators.
- Error control.** The term for general procedures that limit the effects of errors on information represented in noisy, physical systems.
- Error correction.** The process of removing the effects of errors on encoded information.
- Error-correcting code.** A code with additional properties that enable a decoding procedure to remove the effects of the dominant sources of errors on encoded information. Any code is error correcting for some error model in this sense. To call a code error correcting emphasizes the fact that it was designed for this purpose.
- Error model.** An explicit description of how and when errors happen in a given system. Typically, a model is specified as a probability distribution over error operators. More general models may need to be considered, particularly in the context of fault-tolerant computation, for which correlations in time are important.
- Fault tolerance.** A property of encoded information that is being processed with gates. It means that errors occurring during processing, including control errors and environmental noise, do not seriously affect the information of interest.
- Gate.** An operation applied to information for the purpose of information processing.
- Hamming distance.** The Hamming distance between two binary words (sequences of 0 and 1) is the number of positions in which the two words disagree.
- Hilbert space.** An n -dimensional Hilbert space consists of all complex n -dimensional vectors. A defining operation in a Hilbert space is the inner product. If the vectors are thought of as column vectors, then the inner product $\langle x, y \rangle$ of x and y is obtained by forming the conjugate transpose x^\dagger of x and calculating $\langle x, y \rangle = x^\dagger y$. The inner product induces the usual norm $|x|^2 = \langle x, x \rangle$.
- Information.** Something that can be recorded, communicated, and computed with. Information is fungible, which implies that its meaning can be identified regardless of the particulars of the physical realization. Thus, information in one realization (such as ink on a sheet of paper) can be easily transferred to another (for example, spoken words). Types of information include deterministic, probabilistic, and quantum information. Each type is characterized by information units, which are abstract systems whose states represent the simplest information of this type. These define the natural representation of the information. For deterministic information, the unit is the bit, whose states are symbolized by 0 and 1. Information units can be put together to form larger systems and can be processed with basic operations acting on a small number of units at a time.
- Length.** For codes on n basic information units, the length of the code is n .
- Minimum distance.** The smallest number of errors that is not detectable by a code. In this context, the error model consists of a set of error operators without specified probabilities. Typically, the concept is used for codes on n information units, and the error model consists of operators acting on any one of the units. For a classical binary code, the minimum distance is the smallest Hamming distance between two code words.
- Noise.** Any unintended effect on the state of a system, particularly an effect with a stochastic component due to incomplete isolation of the system from its environment.
- Operator.** A function transforming the states of a system. Operators may be restricted, depending on the system's properties. For example, operators acting on quantum systems are always assumed to be linear.
- Pauli operators.** The Hermitian matrices σ_x , σ_y , and σ_z —refer to Equation (7)—acting on qubits. It is often convenient to consider the identity operator to be included in the set of Pauli operators.

Physical system. A system explicitly associated with a physical device or particle.

The term is used to distinguish between abstract systems used to define a type of information and specific realizations, which are subject to environmental noise and errors due to other imperfections.

Probabilistic bit. The basic unit of probabilistic information. It is a system whose state space consists of all probability distributions over the two states of a bit. The states can be thought of as describing the outcome of a biased coin flip before the coin is flipped.

Probabilistic information. The type of information obtained when the state spaces of deterministic information are extended with arbitrary probability distributions over the deterministic states. This is the main type of classical information with which quantum information is compared.

Quantum information. The type of information obtained when the state space of deterministic information is extended with arbitrary superpositions of deterministic states. Formally, each deterministic state is identified with one of an orthonormal basis vector in a Hilbert space, and superpositions are unit-length vectors that are expressible as complex linear sums of the chosen basis vectors. Ultimately, it is convenient to extend this state space again by permitting probability distributions over the quantum states. This is still called quantum information.

Qubit. The basic unit of quantum information. It is the quantum extension of the deterministic bit; that is, its state space consists of the unit-length vectors in a two-dimensional Hilbert space.

Repetition code. The classical, binary repetition code of length n consists of the two words $00 \dots 0$ and $11 \dots 1$. For quantum variants of this code, one applies the superposition principle to obtain the states consisting of all unit-length complex linear combinations of the two classical code words.

Scalability. A property of physical implementations of information processing that implies that there are no bounds on accurate information processing. That is, arbitrarily many information units can be realized, and they can be manipulated for an arbitrarily long amount of time without loss of accuracy. Furthermore, the realization is polynomially efficient in terms of the number of information units and gates used.

States. The set of states for a system characterizes the system's behavior and possible configurations.

Subspace. For a Hilbert space, a subspace is a linearly closed subset of the vector space. The term can be used more generally for a system Q of any information type: A subspace of Q or, more specifically, of the state space of Q is a subset of the state space that preserves the properties of the information type represented by Q .

Subsystem. A typical example of a subsystem is the first (qu)bit in a system consisting of two (qu)bits. In general, to obtain a subsystem of system Q , one first selects a subset C of Q 's state space and then identifies C as the state space of a pair of systems. Each member of the pair is then a subsystem of Q . Restrictions apply, depending on the types of information carried by the system and subsystems. For example, if Q is quantum and so are the subsystems, then C has to be a linear subspace and the identification of the subsystems' state space with C has to be unitary.

Subsystem identification. The mapping or transformation that identifies the state space of two systems with a subset C of states of a system Q . In saying that L is a subsystem of Q , we also introduce a second subsystem and identify the state space of the combined system with the subset of states C .

Syndrome. One of the states of a syndrome subsystem. It is often used more narrowly for one of a distinguished set of basis states of a syndrome subsystem.

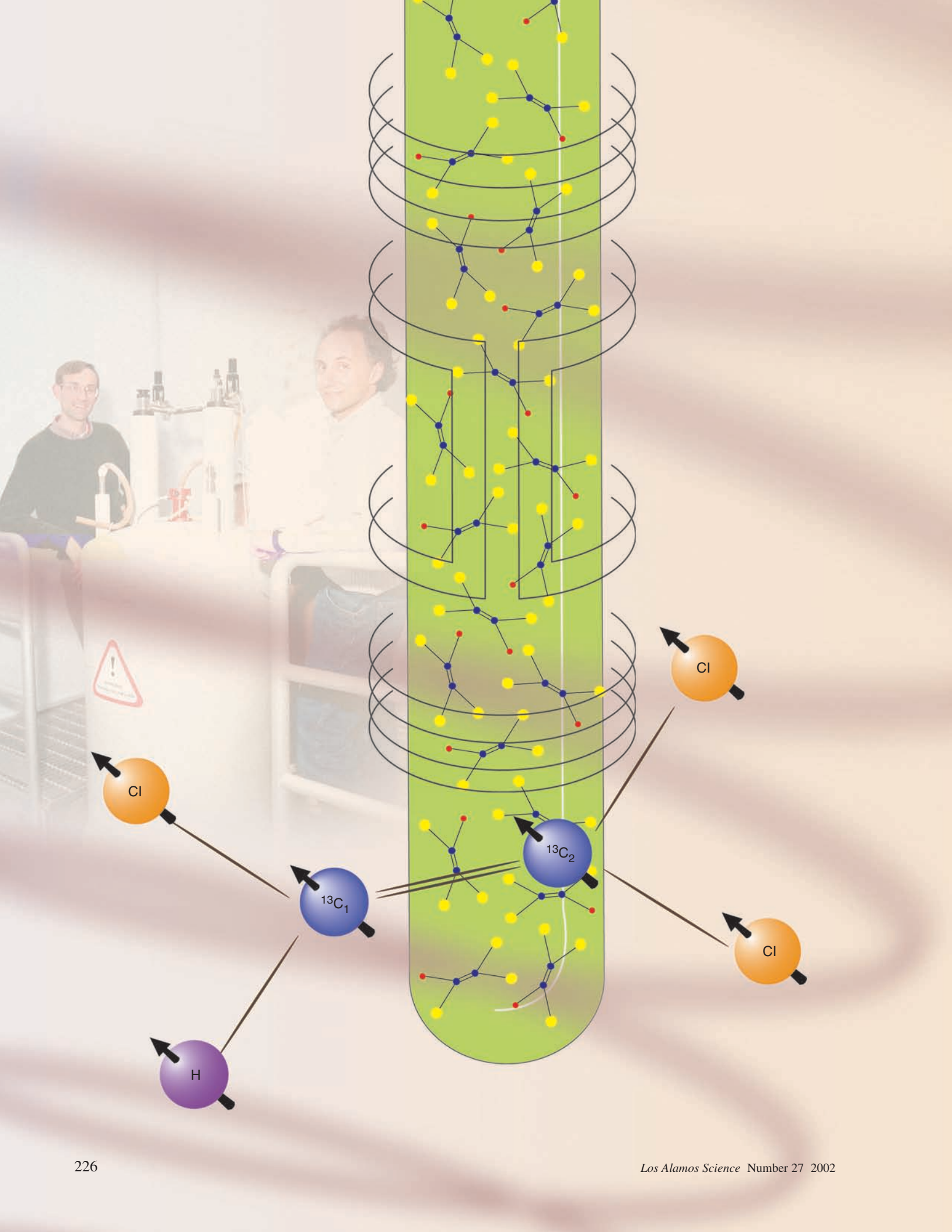
Syndrome subsystem. In identifying an information-carrying subsystem in the context of error correction, the other member of the pair of subsystems required for the subsystem identification is called the syndrome subsystem. The terminology comes from classical error correction, in which the syndrome is used to determine the most likely error that has occurred.

System. An entity that can be in any of a specified number of states. An example is a desktop computer whose states are determined by the contents of its various memories and disks. Another example is a qubit, which can be thought of as a particle whose state space is identified with complex, two-dimensional length-one vectors. Here, a system is always associated with a type of information, which in turn determines the properties of the state space. For example, for quantum information, the state space is a Hilbert space. For deterministic information, it is a finite set called an alphabet.

Twirling. A randomization method for ensuring that errors act like a depolarizing error model. For one qubit, it involves applying a random Pauli operator before the errors occur and then undoing the operator by applying its inverse.

Unitary operator. A linear operator U on a Hilbert space that preserves the inner product. That is, for all x and y , $\langle Ux, Uy \rangle = \langle x, y \rangle$. If U is given in matrix form, then this condition is equivalent to $U^\dagger U = \mathbb{1}$.

Weight. For a binary word, the weight is the number of 1s in the word. For an error operator acting on n systems by applying an operator to each one of them, the weight is the number of nonidentity operators applied.



NMR

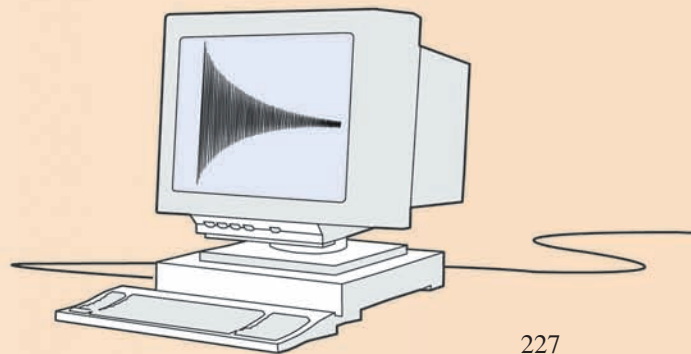
and Quantum Information Processing

Raymond Laflamme, Emanuel Knill, David G. Cory, Evan M. Fortunato, Timothy F. Havel, Cesar Miquel, Rudy Martinez, Camille J. Negrevergne, Gerardo Ortiz, Marco A. Pravia, Yehuda Sharf, Siddhasattwa Sinha, Rolando Somma, and Lorenza Viola

Using quantum physics to represent and manipulate information makes possible surprising improvements in the efficiency with which some problems can be solved. But can these improvements be realized experimentally? If we consider the history of implementing theoretical ideas about classical information and computation, we find that, initially, small numbers of simple devices were used to explore the advantages and difficulties of information processing. For example, in 1933, Atanasoff and his colleagues at the Iowa State College were able to implement digital calculations using about 300 vacuum tubes (Zalta 2002). Although the device was never practical because its error rate was too large, it was probably the first instance of a programmable computer using vacuum tubes, and it opened the way for more stable and reliable devices. Progress toward implementing quantum information processors is also initially confined to limited capacity and error-prone devices.

There are numerous proposals for implementing quantum information processing (QIP) prototypes. To date, however, only three of them have been used to successfully manipulate more than one qubit: cavity quantum electrodynamics (cavity QED), ion traps, and nuclear magnetic resonance (NMR) with molecules in a liquid (or liquid-state NMR). QIP devices are difficult to realize because of an intrinsic conflict between two of the most important requirements: On the one hand, it is necessary for the device to be well isolated from, and therefore interact only weakly with, its environment; otherwise, the crucial quantum correlations on which the advantages of QIP are based are destroyed. On the other hand, it is necessary for the different parts of the device to interact strongly with each other and for some of them to be coupled strongly with the measuring device, which is needed to read out “answers.” That few physical systems have these properties naturally is apparent from the absence of obvious quantum effects in the macroscopic world.

One system whose properties constitute a reasonable compromise between the two requirements consists of the nuclear spins in a molecule in the liquid state. The spins, particularly those with spin $1/2$, provide a natural representation of quantum bits.



They interact weakly but reliably with each other, and the effects of the environment are often small enough. The spins can be controlled with radio-frequency (rf) pulses and observed with measurements of the magnetic fields they generate. Liquid-state NMR has so far been used to demonstrate control of up to seven physical qubits.

It is important to remember that the idea of QIP is less than two decades old, and, with the notable exception of quantum cryptography, experimental proposals and efforts aimed at realizing modern QIP began only in the last five years of the 20th century. Increasingly advanced experiments are being implemented. But from an information processing point of view, we are a long way from using quantum technology to solve an independently posed problem not solvable on a standard personal computer—a typical classical computer. In order to get to the point where such problems can be solved by QIP, current experimental efforts are devoted to understanding the behavior of and the methods for controlling various quantum systems, as well as ways of overcoming their limitations. The work on NMR QIP has focused on the control of quantum systems by algorithmically implementing quantum transformations as precisely as possible. Within the limitations of the device, this approach has been surprisingly successful—thanks to the many scientists and engineers who have perfected NMR spectrometers over the past 50 years.

After a general introduction to NMR, we give the basics of implementing quantum algorithms. We describe how qubits are realized and controlled with rf pulses, their internal interactions, and gradient fields. A peculiarity of NMR is that the internal interactions (given by the internal Hamiltonian) are always on. We discuss how they can be effectively turned off with the help of a standard NMR method called refocusing. Liquid-state NMR experiments are done at room temperature, leading to an extremely mixed (that is, nearly random) initial state. Despite this high degree of randomness, it is possible to investigate QIP because the relaxation time (the time scale over which useful signal from a computation is lost) is sufficiently long. We explain how this feature leads to the crucial ability of simulating a pure (nonrandom) state by using pseudopure states. We discuss how the answer provided by a computation is obtained by measurement and how this measurement differs from the ideal, projective measurement of QIP. We then give implementations of some simple quantum algorithms with a typical experimental result. We conclude with a discussion of what we have learned from NMR QIP so far and what the prospects are for future NMR QIP experiments. For an elementary, device-independent introduction to quantum information and definitions of the states and operators used here, see the article “Quantum Information Processing” on page 2 .

Liquid-State NMR

NMR Basics. Many atomic nuclei have a magnetic moment, which means that, like small bar magnets, they respond to and can be detected by their magnetic fields. Although single nuclei are impossible to detect directly by these means with currently available technology, if sufficiently many are available so that their contributions to the magnetic field add, they can be observed as an ensemble. In liquid-state NMR, the nuclei belong to atoms forming a molecule, a very large number of which are dissolved in a liquid. An example is carbon-13-labeled trichloroethylene (TCE)—see Figure 1. The hydrogen nucleus (that is, the proton) of each TCE molecule has a relatively strong magnetic moment. When the sample is placed in a powerful external magnetic field, each proton’s spin prefers to align itself with the field. It is possible to induce the spin direction to tip off-axis by means of rf pulses, at which point the effect of the static field is to induce a rapid precession of the proton spins. In this introduction, precession refers to a rotation of a spin direction around the main axis, here the z -axis, as determined by

the external magnetic field. The precession frequency ω is often called the Larmor frequency and is linearly related to the strength B of the external field: $\omega = \mu B$, where μ is the magnetic moment. For the proton, the magnetic moment is 42.7 megahertz per tesla (MHz/T), so at a typical field of $B = 11.7$ tesla, the precession frequency is 500 megahertz. The magnetic field produced by the precessing protons induces oscillating currents in a coil judiciously placed around the sample and “tuned” to the precession frequency, allowing observation of the entire ensemble of protons by magnetic induction. This is the fundamental idea of NMR. The device that applies the static magnetic field and rf control pulses and that detects the magnetic induction is called an NMR spectrometer—see Figure 2.

Magnetic induction by nuclear spins was observed for the first time by Edward Purcell and coworkers (1946) and Felix Bloch (1946). This achievement opened a new field of research, leading to many important applications, such as molecular structure determination, dynamics studies both in the liquid and solid state (Ernst et al. 1994), and magnetic resonance imaging (Mansfield and Morris 1982). The application of NMR to QIP is related to methods for determining molecular structure by NMR. Many of the same techniques are used in QIP, but instead of using uncharacterized molecules, specific ones with well-defined nuclear spins are synthesized. In this setting, one can manipulate the nuclear spins as quantum information so that it becomes possible to experimentally demonstrate the fundamental ideas of QIP.

Perhaps the clearest example of early connections of NMR to information theory is the spin echo phenomenon (Hahn 1950). When the static magnetic field is not homogeneous (that is, it is not constant across the sample), the spins precess at different frequencies, depending on their location in the sample. As a result, the magnetic induction signal rapidly vanishes because the magnetic fields produced by the spins are no longer aligned and therefore do not add. The spin echo is used to refocus this effect by inverting the spins, an action that effectively reverses their precession until they are all aligned again. Based on spin echoes, the idea of using nuclear spins for (classical) information storage was suggested and patented by Arthur Anderson et al. (1955) and Anderson and Erwin Hahn (1955).

NMR spectroscopy would not be possible if it were not for relatively long “relaxation” times. Relaxation is the process that tends to realign the nuclear spins with the field and randomize their phases, an effect that leads to complete loss of the information represented in such a spin. In liquid state, relaxation times of the order of seconds are common and attributed to the weakness of nuclear interactions and a fast averaging effect associated with the rapid, tumbling motions of molecules in the liquid state.

Currently, off-the-shelf NMR spectrometers are robust and straightforward to use. The requisite control is to a large extent computerized, so most NMR experiments involve few custom adjustments after the sample has been obtained. Given that the underlying nature of the nuclear spins is intrinsically quantum mechanical, it is not surprising that, soon after Shor’s discovery of the quantum factoring algorithm, NMR was studied as a potentially useful device for QIP.

A Brief Survey of NMR QIP. Concrete and workable proposals for using liquid-state NMR for quantum information were first given by David Cory et al. (1997) and Neil Gershenfeld and Isaac Chuang (1997). Three difficulties had to be overcome for NMR QIP to become possible. The first was that the standard definitions of quantum information and computation require that quantum information be stored in a single physical system. In NMR, an obvious such system consists of some of the nuclear spins

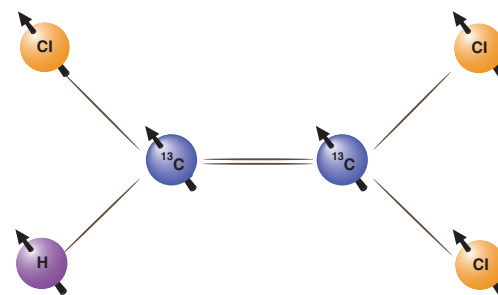


Figure 1. Schematic of a Typical Molecule (Trichloroethylene) Used for QIP

There are three useful nuclei for realizing qubits. They are the proton (H) and the two carbons (^{13}C). The molecule is “labeled,” which means that the nuclei are carefully chosen isotopes. In this case, the normally predominant isotope of carbon, ^{12}C (a spin-zero nucleus), is replaced by ^{13}C , which has spin 1/2.

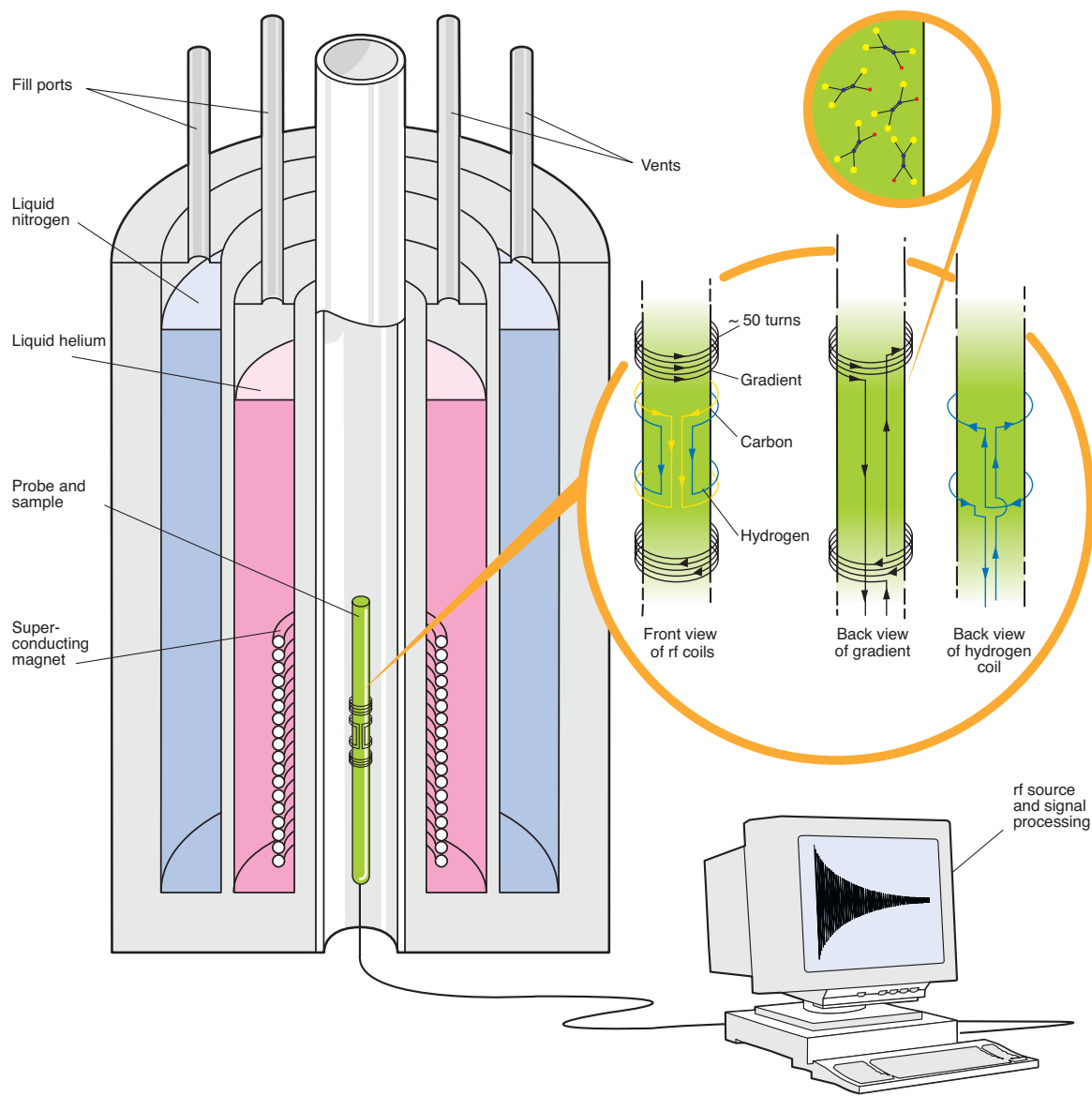


Figure 2. Schematic of a Typical NMR Spectrometer (not to scale)

The main components of a spectrometer are the magnet, which is superconducting, and the console, which has the electronics needed to control the spectrometer. The sample containing a liquid solution of the molecule used for QIP is inserted into the central core of the magnet, where it is surrounded by the probe. The probe (shown enlarged to the right) contains coils for applying the rf pulses and magnetic field gradients.

in a single molecule. But it is not possible to detect single molecules with available NMR technology. The solution that makes NMR QIP possible can be applied to other QIP technologies: Consider the large collection of available molecules as an ensemble of identical systems. As long as they all perform the same task, the desired answers can be read out collectively. The second difficulty was that the standard definitions require that readout take place by a projective quantum measurement of the qubits. From such a measurement, one learns whether a qubit is in the state $|0\rangle$ or $|1\rangle$. The two measurement outcomes have probabilities determined by the initial state of the qubits being used, and after the measurement, the state collapses to a state consistent with the outcome.

The measurement in NMR is much too weak to determine the outcome and cause the state's collapse for each molecule. But because of the additive effects of the ensemble, one can observe a (noisy) signal that represents the average, over all the molecules, of the probability that $|1\rangle$ would be the outcome of a projective measurement. It turns out that this so-called weak measurement suffices for realizing most quantum algorithms, in particular those whose ultimate answer is deterministic. Shor's factoring and Lov Grover's search algorithms can be modified to satisfy this property. The final and most severe difficulty was that, even though in equilibrium there is a tendency for the spins to align with the magnetic field, the energy associated with this tendency is very small compared with room temperature. Therefore, the equilibrium states of the molecules' nuclear spins are nearly random, with only a small fraction pointing in the right direction. This difficulty was overcome by methods for singling out the small fraction of the observable signal that represents the desired initial state. These methods were anticipated in 1977 (Stall et al.)

Soon after these difficulties were shown to be overcome or circumventable, two groups were able to experimentally implement short quantum algorithms using NMR with small molecules (Chuang et al. 1998, Jones et al. 1998). At present, it is considered unlikely that liquid-state NMR algorithms will solve problems not easily solvable with available classical computing resources. Nevertheless, experiments in liquid-state NMR QIP are remarkable for demonstrating that one can control the unitary evolution of physical qubits sufficiently well to implement simple QIP tasks. The control methods borrowed from NMR and developed for the more complex experiments in NMR QIP are applicable to other device technologies, enabling better control in general.

Principles of Liquid-State NMR QIP

In order to physically realize quantum information, it is necessary to find ways of representing, manipulating, and coupling qubits so as to implement nontrivial quantum gates, prepare a useful initial state, and read out the answer. The next sections show how to accomplish these tasks in liquid-state NMR.

Realizing Qubits. The first step for implementing QIP is to have a physical system that can carry quantum information. The preferred system for realizing qubits in liquid-state NMR consists of spin-1/2 nuclei, which are naturally equivalent to qubits. The nuclear-spin degree of freedom of a spin-1/2 nucleus defines a quantum mechanical two-state system. Once the direction along the strong external magnetic field is fixed, its state space consists of the superpositions of "up" and "down" states. That is, we can imagine that the nucleus behaves somewhat like a small magnet, with a definite axis, which can point either up (logical state $|0\rangle$) or down (logical state $|1\rangle$). By the superposition principle, every quantum state of the form $|\psi_0\rangle = \alpha|0\rangle + \beta|1\rangle$ with $|\alpha|^2 + |\beta|^2 = 1$ is a possible (pure) state for the nuclear spin. In the external magnetic field, the two logical states have different energies. In quantum mechanics, this observation means that the time evolution of $|\psi_0\rangle$ is given by

$$|\psi_t\rangle = e^{-i\omega t/2}\alpha|0\rangle + e^{i\omega t/2}\beta|1\rangle . \quad (1)$$

The constant ω is the precession frequency of the nuclear spin in the external magnetic field in units of radians per second if t is in seconds. The frequency is proportional to the energy difference ε between the up and down states: $\omega = 2\pi\varepsilon/h$, where h is Planck's constant.

Although a spin-1/2 nucleus' state space is the same as that of a qubit, the precession implies that the state is not constant. We would like the realization of a qubit to retain its state over time when we are not intentionally modifying it. For this reason, in the next section, the qubit state realized by the nuclear spin will be defined so as to compensate for the precession.

Precession frequencies for nuclear spins can vary substantially depending on the nuclei's magnetic moments. For example, at 11.7 tesla, the precession frequency for protons is 500 megahertz, and for carbon-13, it is 125 megahertz. These frequency differences are exploited in measurement and control to distinguish between the types of nuclei. The effective magnetic field seen by nuclear spins also depends on their chemical environment. This dependence causes small variations in the spins' precession frequencies that can be used to distinguish, for example, the two carbon-13 nuclei in TCE: The frequency difference (called the "chemical shift") is 600 to 900 hertz at 11.7 tesla, depending on the solvent, the temperature, and the TCE concentration.

Pauli Matrices

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

If we use the Pauli matrix σ_z , the time evolution can be expressed as $|\psi_t\rangle = e^{i\omega\sigma_z t/2}|\psi_0\rangle$. The operator $\omega\sigma_z/2$ is the internal Hamiltonian (that is, the energy observable, in units for which $\hbar/(2\pi) = 1$) of the nuclear spin. The direction of the external magnetic field determines the z-axis. Given a choice of axes, the idea that a single nuclear spin 1/2 has a direction (as would be expected for a tiny magnet) can be made explicit by means of the Bloch sphere representation of a nuclear spin's state (refer to Figure 3). The Pauli matrix σ_z can be thought of as the observable that measures the nuclear spin along the z-axis. Observables for spin along the x- and y-axis are given by the other two Pauli matrices, σ_x and σ_y . Given a state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ of the nuclear spin, one can form the density matrix $|\psi\rangle\langle\psi|$ and express it in the form

$$|\psi\rangle\langle\psi| = \frac{1}{2}(\mathbb{1} + \alpha_x\sigma_x + \alpha_y\sigma_y + \alpha_z\sigma_z) . \tag{2}$$

The vector $\mathbf{v} = (\alpha_x, \alpha_y, \alpha_z)$ then is a point on the unit sphere in three-dimensional space. Conversely, every point on the unit sphere corresponds to a pure state of the nuclear spin. The representation also works for mixed states, which correspond to points in the interior of the sphere. As a representation of spin states, the unit sphere is called the Bloch sphere. Because quantum evolutions of a spin correspond to Bloch sphere rotations, the Bloch sphere is a useful tool for thinking about one- and sometimes about two-qubit processes.

If we write the state as a density matrix ρ and expand it in terms of Pauli matrices,

$$\begin{aligned} \rho &= |\psi\rangle\langle\psi| = (\mathbb{1} + x\sigma_x + y\sigma_y + z\sigma_z)/2 \\ &= \frac{1}{2}(\mathbb{1} + \sin(\theta)\cos(\phi)\sigma_x + \sin(\theta)\sin(\phi)\sigma_y + \cos(\theta)\sigma_z) , \end{aligned} \tag{3}$$

the coefficients $(x, y, z) = (\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta))$ of the Pauli matrices form the vector for the state. The angles θ and ϕ are the Euler angles, as shown in Figure 3. For a pure state, this vector is on the surface of the unit sphere, and for a mixed state, it is inside the unit sphere. The Pauli matrices are associated with spin observables in the laboratory frame, so that all axes of the representation are meaningful with respect to real space.

One-Qubit Gates. The second step for realizing QIP is to give a means for controlling the qubits so that quantum algorithms can be implemented. The qubits are controlled with carefully modulated external fields to realize specific unitary evolutions called gates.

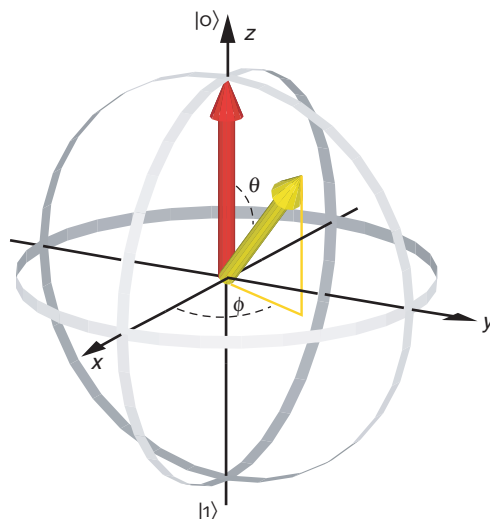


Figure 3. Bloch Sphere Representation of a Qubit State

The yellow arrow represents a pure state $|\psi\rangle$ for the qubit or the nuclear spin $1/2$. The Euler angles are indicated and determine the state according to the formula $|\psi\rangle = \cos(\theta/2)|0\rangle + e^{i\phi} \sin(\theta/2)|1\rangle$. The red arrow along the z-axis indicates the orientation of the magnetic field and the vector for $|0\rangle$.

Each such evolution can be described by a unitary operator applied to one or more qubits. The simplest method for demonstrating that sufficient control is available is to show how to realize a set of one- and two-qubit gates that is universal in the sense that, in principle, every unitary operator can be implemented as a composition of gates (Barenco et al. 1995, DiVincenzo 1995, Lloyd 1995).

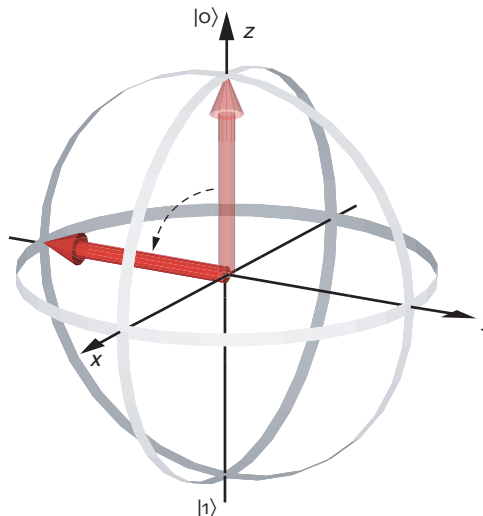
One-qubit gates can be thought of as rotations of the Bloch sphere and can be implemented in NMR with electromagnetic pulses. In general, the effect of a magnetic field on a nuclear spin is to cause a rotation around the direction of the field. In terms of the quantum state of the spin, the effect is described by an internal Hamiltonian of the form $H = (\omega_x \sigma_x + \omega_y \sigma_y + \omega_z \sigma_z)/2$. The coefficients of the Pauli matrices depend on the magnetic field according to $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z) = -\mu \mathbf{B}$, where μ is the nuclear magnetic moment and \mathbf{B} is the magnetic field vector. In terms of the Hamiltonian, the evolution of the spin's quantum state in the presence of the magnetic field \mathbf{B} is therefore given by $|\psi_t\rangle = e^{-iHt}|\psi_0\rangle$ so that the spin direction in the Bloch sphere rotates around $\boldsymbol{\omega}$ with angular frequency $\omega = |\boldsymbol{\omega}|$.

In the case of liquid-state NMR, there is an external, strong magnetic field along the z-axis, and the applied electromagnetic pulses add to this field. One can think of these pulses as contributing a relatively weak magnetic field (typically less than .001 of the external field), whose orientation is in the xy-plane. One use of such a pulse is to tip the nuclear spin from the z-axis to the xy-plane. To see how that can be done, assume that the spin starts in the state $|0\rangle$, which points up along the z-axis in the Bloch sphere representation. Because this state is aligned with the external field, it does not precess. To tip the spin, one can start by applying a pulse field along the x-axis. Because the pulse field is weak compared with the external field, the net field is still almost along the z-axis. The spin now rotates around the net field. Since it started along z, it moves only in a small circle near the z-axis. To force the spin to tip further, one changes the orientation of the pulse field at the same frequency as the precession frequency. This is called a resonant pulse. Because typical precession frequencies are hundreds of megahertz, such a pulse consists of rf electromagnetic fields.

To better understand how resonant pulses work, it is convenient to use the "rotating frame." In this frame, we imagine that our apparatus rotates at the precession frequency of the nuclear spin. In this way, the effect of the external field is removed. In particular, in the rotating frame, the nuclear spin does not precess, and a resonant pulse's magnetic field looks like a constant magnetic field applied, for example, along the $(-x)$ -axis of the rotating frame. The nuclear spin responds to the pulse by rotating around the x-axis, as expected: If the spin starts along the z-axis, it tips toward the $(-y)$ -axis, then goes to

Figure 4. Single-Bit Rotation around the x -Axis in the Rotating Frame

An applied magnetic field along the rotating frame's ($-x$)-axis due to a resonant rf pulse moves the nuclear spin direction from the z - toward the ($-y$)-axis. The initial and final states for the nuclear spin are shown for a 90° rotation. If the strength of the applied magnetic field is such that the spin evolves according to the Hamiltonian $\omega_x \sigma_x / 2$, then it has to be turned on for a time $t = \pi / (2\omega_x)$ to cause the rotation shown.



the ($-z$)-, to the y -, and finally back to the z -axis, all in the rotating frame (see Figure 4).

The rotating frame makes it possible to define the state of the qubit realized by a nuclear spin as the state with respect to this frame. As a result, the qubit's state does not change unless rf pulses are applied. In the context of the qubit realized by a nuclear spin, the rotating frame is called the logical frame. In the following, references to the Bloch sphere axes and associated observables are understood to be with respect to an appropriate, usually rotating, frame. Different frames can be chosen for each nuclear spin of interest, so we often use multiple independently rotating frames and refer each spin's state to the appropriate frame.

Use of the rotating frame together with rf pulses makes it possible to implement all one-qubit gates on a qubit realized by a spin-1/2 nucleus. To apply a rotation around the x -axis, a resonant rf pulse with effective field along the rotating frame's ($-x$)-axis is applied. This is called an x -pulse, and x is the "axis" of the pulse. While the rf pulse is on, the qubit's state evolves as $e^{-i\omega_x \sigma_x t / 2}$. The strength (or power) of the pulse is characterized by ω_x , the nutation frequency. To implement a rotation by an angle of ϕ , the pulse is turned on for a period $t = \phi / \omega_x$. Rotations around any axis in the plane can be implemented similarly. The angle of the pulse field with respect to the ($-x$)-axis is called the phase of the pulse. It is a fact that all rotations of the Bloch sphere can be decomposed into rotations around axes in the plane. For rotations around the z -axis, an easier technique is possible. The current absolute phase θ of the rotating frame's x -axis is given by $\theta_0 + \omega t$, where ω is the precession frequency of the nuclear spin. Changing the angle θ_0 by $-\phi$ is equivalent to rotating the qubit's state by ϕ around the z -axis. In this sense, z -pulses can be implemented exactly. In practice, this change of the rotating frame's phase means that the absolute phases of future pulses must be shifted accordingly. This implementation of rotations around the z -axis is possible because phase control in modern equipment is extremely reliable so that errors in the phase of applied pulses are negligible compared with other sources of errors.

So far, we have considered just one nuclear spin in a molecule. But the rf fields are experienced by the other nuclear spins as well. This side effect is a problem if only one target nuclear spin's state is to be rotated. There are two cases to consider depending on the precession frequencies of the other, nontarget spins. Spins of nuclei of different isotopes, such as those of other species of atoms, usually have precession frequencies that differ from the target's by many megahertz at 11.7 tesla. A pulse resonant for the target has little effect on such spins because, in the rotating frames of the nontarget spins, the pulse's magnetic field is not constant but rotates rapidly. The power of a typical pulse is such that the effect during one rotation of the pulse's field direction is insignificant and

averages to zero over many rotations. This is not the case for nontarget spins of the same isotope. Although the variations in their chemical environments result in frequency differences, these differences are much smaller, often only a few kilohertz. The period of a 1-kilohertz rotation is 1 millisecond, whereas so-called hard rf pulses require only tens of microseconds (.001 millisecond) to complete the typical 90° or 180° rotations. Consequently, in the rotating frame of a nontarget spin with a small frequency difference, a hard rf pulse's magnetic field is nearly constant for the duration of the pulse. As a result, such a spin experiences a rotation similar to the one intended for the target. To rotate a specific nuclear spin or spins within a narrow range of precession frequencies, one can use weaker, longer-lasting "soft" pulses instead. This approach leads to the following strategies for applying pulses: To rotate all the nuclear spins of a given species (such as the two carbon-13 nuclei of TCE) by a desired angle, apply a hard rf pulse for as short a time as possible. To rotate just one spin having a distinct precession frequency, apply a soft rf pulse of sufficient duration to have little effect on other spins. The power of soft pulses is usually modulated in time ("shaped") to reduce the time needed for a rotation while minimizing crosstalk, a term that describes unintended effects on other nuclear spins.

Two-Qubit Gates. Two nuclear spins in a molecule interact with each other, as one would expect of two magnets. But the details of the spins' interaction are more complicated because they are mediated by the electrons. In liquid state, the interaction is also modulated by the rapid motions of the molecule. The resulting effective interaction is called the J -coupling. When the difference of the precession frequencies between the coupled nuclear spins is large compared with the strength of the coupling, it is a good approximation to write the coupling Hamiltonian as a product of the z -Pauli operators for each spin: $H_J = C\sigma_z^{(1)}\sigma_z^{(2)}$. This is the weak-coupling regime. With this Hamiltonian, an initial state $|\psi_0\rangle$ of two nuclear-spin qubits evolves as $|\psi_t\rangle = e^{-iC\sigma_z^{(1)}\sigma_z^{(2)}t}|\psi_0\rangle$, where a different rotating frame is used for each nuclear spin to eliminate the spin's internal evolution. (The use of rotating frames is compatible with the coupling Hamiltonian because the Hamiltonian is invariant under frame rotations.) Because the Hamiltonian is diagonal in the logical basis, the effect of the coupling can be understood as an increase of the (signed) precession frequency of the second spin if the first one is up and a decrease if the first one is down (see Figure 5). The changes in precession frequency for adjacent nuclear spins in organic molecules are typically in the range of 20 to 200 hertz. They are normally much smaller for non-adjacent nuclear spins. The strength of the coupling is called the coupling constant and is given as the change in the precession frequency. In terms of the constant C used above, the coupling constant is given by $J = 2C/\pi$ in hertz. For example, the coupling constants in TCE are close to 100 hertz between the two carbons, 200 hertz between the proton and the adjacent carbon, and 9 hertz between the proton and the far carbon.

The J -coupling and the one-qubit pulses suffice for realizing the controlled-not operation usually taken as one of the fundamental gates of QIP. A pulse sequence for implementing the controlled-not in terms of the J -coupling constitutes the first quantum algorithm discussed under "Examples of Quantum Algorithms for NMR." A problem with the J -coupling in liquid-state NMR is that it cannot be turned off when it is not needed for implementing a gate.

Turning off the J -Coupling. The coupling between the nuclear spins in a molecule cannot be physically turned off. But for QIP, we need to be able to maintain a state in memory and to couple qubits selectively. Fortunately, NMR spectroscopists solved this problem well before the development of modern quantum-information concepts. The idea is to use the control of single spins to cancel the interaction's effect over a given

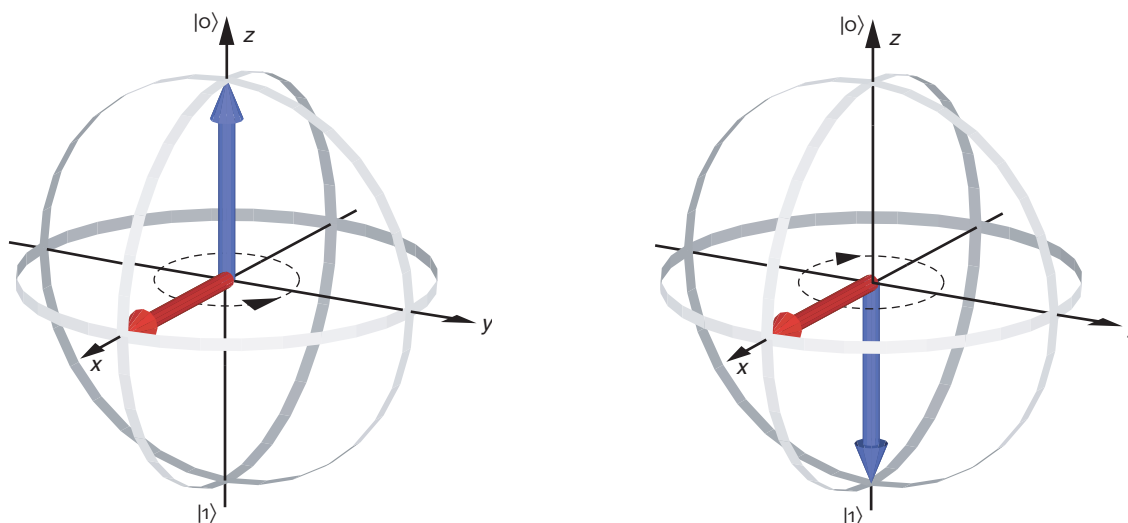


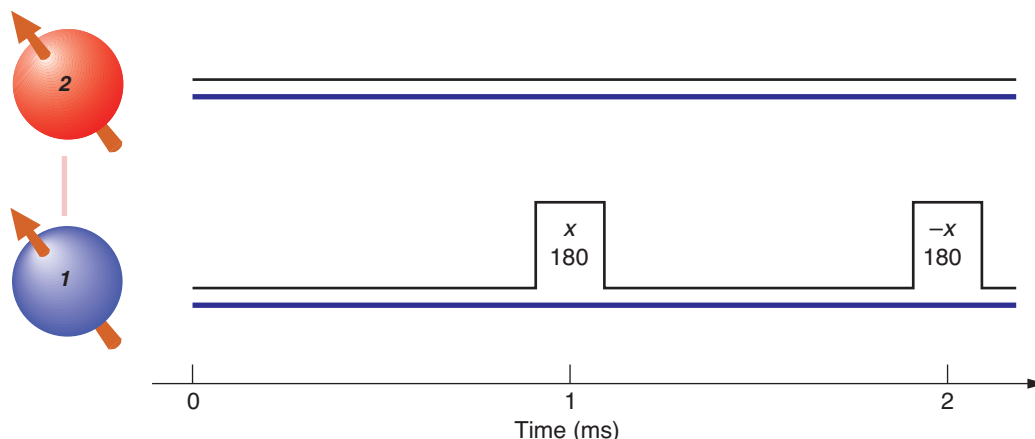
Figure 5. J-Coupling Effect

In the weak-coupling regime with a positive coupling constant, the coupling between two spins can be interpreted as an increase in precession frequency of spin 2 when spin 1 is up and a decrease when spin 1 is down. The two diagrams depict the situation in which spin 2 is in the plane. The diagram on the left has spin 1 pointing up along the z-axis. In the rotating frame of spin 2, it precesses from the x-axis to the y-axis. The diagram on the right has spin 1 pointing down, causing a precession in the opposite direction of spin 2. Note that neither the coupling nor the external field changes the orientation of a spin pointing up or down along the z-axis.

period. This technique is called refocusing and requires applying a 180° pulse to one of two coupled spins at the midpoint of the desired period. To understand how refocusing works, consider again the visualization of Figure 5. A general state is in a superposition of the four logical states of the two spins. By linearity, it suffices to consider the evolution with spin 1 being in one of its two logical states, up or down, along the z-axis. Suppose we wish to remove the effects of the coupling over a period of 2 milliseconds. To do so, wait 1 millisecond. In a sequence of pulses, this waiting period is called a 1-millisecond delay. The effect on spin 2 in its rotating frame is to precess counterclockwise if spin 1 is up and clockwise for the same angle if spin 1 is down. Now, apply a pulse that rotates spin 1 by 180° around the x-axis. This is called an inversion, or in the current context, a refocusing pulse. It exchanges the up and down states. For the next 1 millisecond, the effect of the coupling on spin 2 is to undo the earlier rotation. At the end of the second 1-millisecond delay, one can apply another 180° pulse to reverse the inversion and recover the initial state. The pulse sequence is depicted in Figure 6.

Turning off couplings between more than two nuclear spins can be quite complicated unless one takes advantage of the fact that nonadjacent nuclear spins tend to be relatively weakly coupled. Methods that scale polynomially with the number of nuclear spins and that can be used to selectively couple pairs of nuclear spins can be found in Debbie Leung et al. (1999) and Jonathan Jones and Knill (1999). These techniques can be used in other physical systems, where couplings exist that are difficult to turn off directly. An example is qubits represented by the state of one or more electrons in tightly packed quantum dots.

Measurement. To determine the answer of a quantum computation, it is necessary to make a measurement. As noted earlier, the technology for making a projective measurement of individual nuclear spins does not yet exist. In liquid-state NMR, instead of using just one molecule to define a single quantum register, we use a large ensemble of molecules in a test tube. Ideally, their nuclear spins are all placed in the same initial state, and the subsequent rf pulses affect each molecule in the same way. As a result, weak magnetic signals from, say, the proton spins in TCE add to form a detectable magnetic field called the bulk magnetization. The signal that is measured in high-field NMR is the magnetization in the xy-plane, which can be picked up by coils whose axes are placed transversely to the external field. Because the interaction of any given nuclear spin with the coil is very weak, the effect of the coil on the quantum state of the spins is negligible in most NMR experiments. As a result, it is a good approximation to think of the generated magnetic fields and their detection classically. In this approximation, each nuclear spin behaves like a tiny bar magnet and contributes to the bulk magnetization.



As the nuclear spins precess, so does the magnetization. As a result, an oscillating current is induced in the coil, provided it is electronically configured to be tuned to the precession frequency. By observing the amplitude and phase of this current over time, we can keep track of the absolute magnetization in the plane and its phase with respect to the rotating frame. This process yields information about the qubit states represented by the state of the nuclear spins.

To see how one can use bulk magnetization to learn about the qubit states, consider the TCE molecule with three spin-1/2 nuclei used for information processing. The bulk magnetizations generated by the protons and the carbons precess at 500 megahertz and 125 megahertz, respectively. The proton and carbon contributions to the magnetization are detected separately with two coils tuned to 500 megahertz (proton magnetization) and 125 megahertz (carbon magnetization). For simplicity, we restrict our attention to the two carbons and assume that the protons are not interacting with the carbons. (It is possible to actively remove such interactions by using a technique called decoupling.)

At the end of a computation, the qubit state of the two nuclear spins is given by a density matrix ρ_q . We can assume that this state is the same for each TCE molecule in the sample. As we mentioned earlier, the density matrix is relative to logical frames for each nuclear spin. The current phases for the two logical frames with respect to a rotating reference frame at the precession frequency of the first carbon are known. If we learn something about the state in the reference frame, that information can be converted to the desired logical frame by a rotation around the z -axis. Let $\rho(0)$ be the state of the two nuclear spins in the reference frame. In this frame, the state evolves in time as $\rho(t)$ according to a Hamiltonian H that consists of a chemical shift term for the difference in the precession frequency of the second carbon and of a coupling term. To a good approximation,

$$H = \pi 900\text{Hz} \sigma_z^{(2)} + \pi 50\text{Hz} \sigma_z^{(1)} \sigma_z^{(2)}. \quad (4)$$

The magnetization detected in the reference x -direction at time t is given by

$$M_x(t) = m \text{tr} \left(\rho(t) \left(\sigma_x^{(1)} + \sigma_x^{(2)} \right) \right), \quad (5)$$

Figure 6. Pulse Sequence for Refocusing the Coupling

The sequence of events is shown with time running from left to right. The two spins' lifelines are shown in blue, and the rf power targeted at each spin is indicated by the black line above. Pulses are applied to spin 1 only, as indicated by the rectangular rises in rf power at 1 ms and 2 ms. The axis for each pulse is given with the pulse. The angle is determined by the area under the pulse and is also given explicitly. Ideally, for pulses of this type, the pulse times (the widths of the rectangles) should be zero. In practice, for hard pulses, they can be as small as $\approx .01$ ms. Any $\sigma_z^{(1)} \sigma_z^{(2)}$ coupling's effect is refocused by the sequence shown so that the final state of the two spins is the same as the initial state. The axis for the pair of refocusing pulses can be changed to any other axis in the plane.

where $\text{tr}(\sigma)$ denotes the trace, that is, the sum of the diagonal elements of the matrix σ . Equation 5 links the magnetization to the Bloch sphere representation. The constant of proportionality m depends on the size of the ensemble and the magnetic moments of the nuclei. From the point of view of NMR, m determines a scale whose absolute size is not relevant. What matters is how strong this signal is compared with the noise in the system. For the purpose of the following discussion, we set $m = 1$.

We can also detect the magnetization $M_y(t)$ in the y -direction and use this result together with $M_x(t)$ to form a complex number representing the planar magnetization.

$$M(t) = M_x(t) + iM_y(t) \quad (6)$$

$$= \text{tr} \left(\rho(t) \left(\sigma_+^{(1)} + \sigma_+^{(2)} \right) \right), \quad (7)$$

where we defined $\sigma_+ = \sigma_x + i\sigma_y = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$.

What can we infer about $\rho(0)$ from observing $M(t)$ over time? For the moment, we neglect the coupling Hamiltonian. Under the chemical shift Hamiltonian $H_{CS} = \pi 900 \text{Hz} \sigma_z^{(2)}$, $M(t)$ evolves as

$$\begin{aligned} M(t) &= \text{tr} \left(e^{-iH_{CS}t} \rho(0) e^{iH_{CS}t} \left(\sigma_+^{(1)} + \sigma_+^{(2)} \right) \right) \\ &= \text{tr} \left(\rho(0) e^{iH_{CS}t} \left(\sigma_+^{(1)} + \sigma_+^{(2)} \right) e^{-iH_{CS}t} \right) && \text{Use } \text{tr}(AB) = \text{tr}(BA). \\ &= \text{tr} \left(\rho(0) \left(\sigma_+^{(1)} + e^{iH_{CS}t} \sigma_+^{(2)} e^{-iH_{CS}t} \right) \right) && H_{CS} \text{ acts only on spin 2.} \\ &= \text{tr} \left(\rho(0) \left(\sigma_+^{(1)} + e^{i2\pi 900 \text{Hz} t} \sigma_+^{(2)} \right) \right) && \text{Multiply the matrices.} \\ &= \text{tr} \left(\rho(0) \sigma_+^{(1)} \right) + \text{tr} \left(\rho(0) e^{i2\pi 900 \text{Hz} t} \sigma_+^{(2)} \right). && \text{The trace is linear.} \end{aligned} \quad (8)$$

Thus, the signal is a combination of a constant signal given by the first spin's contribution to the magnetization in the plane and a signal oscillating with a frequency of 900 hertz with amplitude given by the second spin's contribution to the planar magnetization. The two contributions can be separated by Fourier-transforming $M(t)$, which results in two distinct peaks, one at 0 hertz and a second at 900 hertz (refer to Figure 7).

To see how the coupling affects the observed magnetization, we rewrite the expression for $M(t)$ to take advantage of the fact that the up-down states are invariant under the full Hamiltonian.

$$\begin{aligned} M(t) &= \text{tr} \left(\rho(t) \sigma_+^{(1)} \right) + \text{tr} \left(\rho(t) \sigma_+^{(2)} \right) \\ &= \text{tr} \left(\rho(t) \sigma_+^{(1)} \mathbb{1}^{(2)} \right) + \text{tr} \left(\rho(t) \mathbb{1}^{(1)} \sigma_+^{(2)} \right) \\ &= \text{tr} \left(\rho(t) \sigma_+^{(1)} \left(e_{\uparrow}^{(2)} + e_{\downarrow}^{(2)} \right) \right) + \text{tr} \left(\rho(t) \left(e_{\uparrow}^{(1)} + e_{\downarrow}^{(1)} \right) \sigma_+^{(2)} \right), \end{aligned} \quad (9)$$

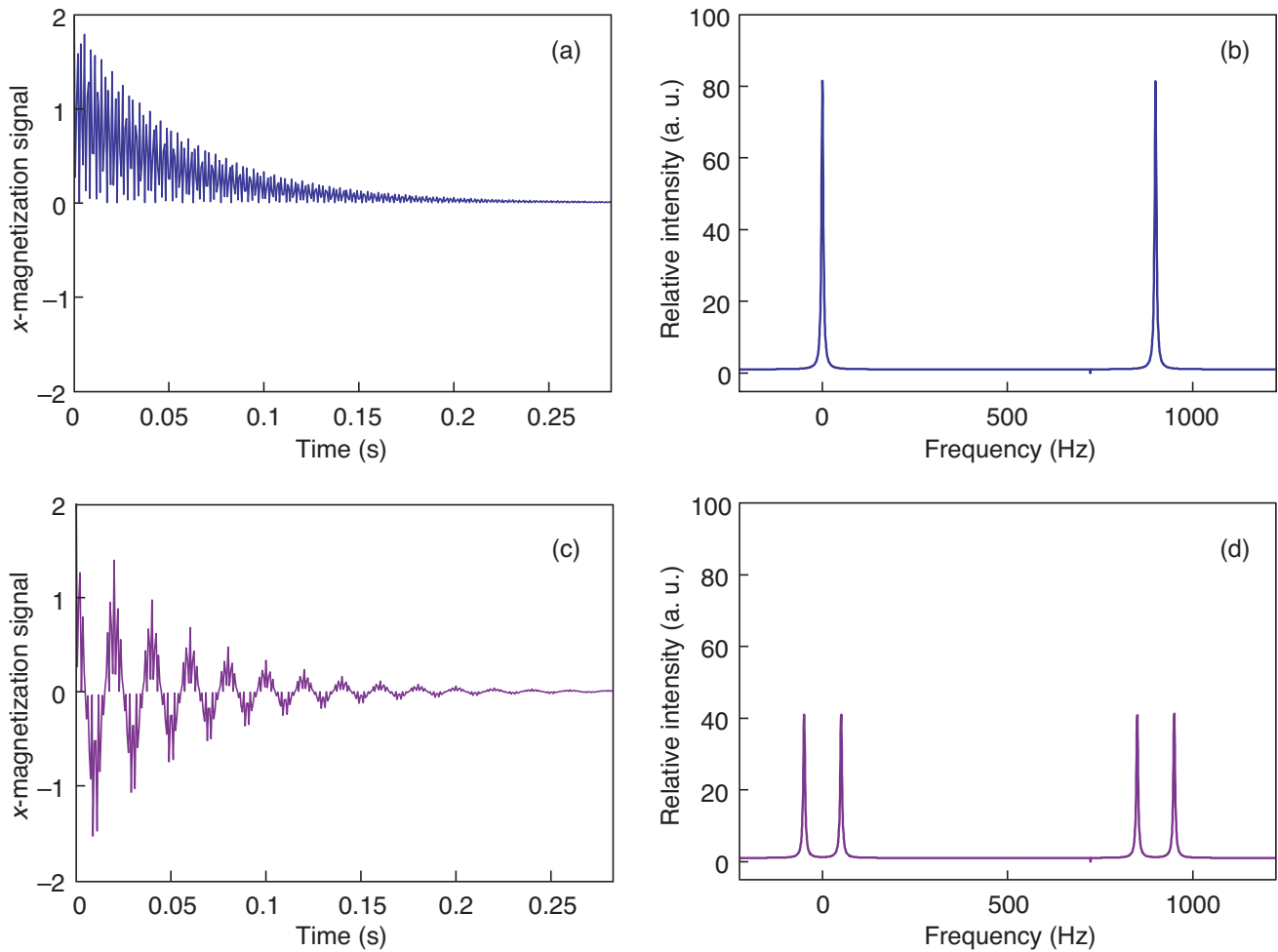


Figure 7. Simulated Magnetization Signals and Spectra

(a) The x -magnetization signal is shown as a function of time for a pair of uncoupled spins with a relative chemical shift of 900 Hz. The initial spin directions are along the x -axis. The signal (called the “free-induction decay”) decays with a halftime of 0.0385 s because of simulated relaxation processes. Typically, the halftimes are much longer. A short one was chosen in order to broaden the peaks for visual effect. (b) The spectrum, that is, the Fourier transform of the combined x - and y -magnetization has peaks at frequencies of 0 Hz (spin 1’s peak) and 900 Hz (spin 2’s peak) because

of the independently precessing pair of spins. (c) This plot shows the x -magnetization signal when the two spins coupled as described in the text. (d) Shown here is the spectrum for the signal in (c) obtained from combined x - and y -magnetization. Each spin’s peak from the previous spectrum “splits” into two. The left and right peaks of each pair are associated with the other spin being in the state $|1\rangle$ and $|0\rangle$, respectively. The vertical axis units are relative intensity with the same constant of proportionality for the two spectra.

where
$$e_{\uparrow} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } e_{\downarrow} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} .$$

Using a calculation similar to the one leading to Equation (8), the first term can be written as

$$M_1(t) = \text{tr} \left(e^{-iHt} \rho(0) e^{iHt} \sigma_+^{(1)} (e_{\uparrow}^{(2)} + e_{\downarrow}^{(2)}) \right) \tag{10}$$

$$= e^{i2\pi 50\text{Hz}t} \text{tr} \left(\rho(0) \sigma_+^{(1)} e_{\uparrow}^{(2)} \right) + e^{-i2\pi 50\text{Hz}t} \text{tr} \left(\rho(0) \sigma_+^{(1)} e_{\downarrow}^{(2)} \right) , \tag{11}$$

and similarly for the second term, but with an offset frequency of 900 hertz because of the chemical shift. It can be seen that the zero-frequency signal splits into two signals with frequencies of -50 hertz and 50 hertz, respectively. The difference between the two frequencies is the coupling constant. The amplitudes of the different frequency signals can be used to infer the expectations of operators such as $\sigma_+^{(1)}e_{\uparrow}^{(2)}$, given by $\text{tr}(\rho(0)\sigma_+^{(1)}e_{\uparrow}^{(2)})$. For n spin-1/2 nuclei, the spectral peak of a nucleus splits into a group of 2^{n-1} peaks, each associated with operators such as $\sigma_+^{(a)}e_{\uparrow}^{(b)}e_{\downarrow}^{(c)}e_{\downarrow}^{(d)}\dots$. Later in the article (see figure on page 249), we show a simulated peak group for a nuclear spin coupled to three other spins. Expectations of the single-spin operators $\sigma_x^{(a)}$ and $\sigma_y^{(a)}$ can be obtained from the real and imaginary parts of the total signal in a peak group for a nucleus. The positions of the 2^{n-1} peaks depend on the couplings. If the peaks are well separated, we can infer expectations of product operators with only one σ_x or σ_y , such as $\sigma_x^{(a)}\sigma_z^{(b)}\mathbb{1}^{(c)}\sigma_z^{(d)}$, by taking linear combinations with appropriate coefficients of the peak amplitudes in a peak group.

In addition to the unitary evolution due to the internal Hamiltonian, relaxation processes tend to decay $\rho(t)$ toward the equilibrium state. In liquid state, the equilibrium state ρ_{thermal} is close to $\mathbb{1}/N$, where N is the total dimension of the state space. The difference between ρ_{thermal} and $\mathbb{1}/N$ is the equilibrium “deviation” density matrix and has magnetization only along the z -axis (see the section “The Initial State”). Because the only observed magnetization is planar, the observed signal decays to zero as the state relaxes to equilibrium. To a good approximation, we can write

$$\rho(t) = \frac{1}{N} \mathbb{1} + e^{-\lambda t} \rho'(t) + (\text{not observed}) , \tag{12}$$

where $\rho'(t)$ has trace zero and evolves unitarily under the Hamiltonian. The effect of the relaxation process is that $M(t)$ has an exponentially decaying envelope, explaining the conventional name for $M(t)$, namely, the free induction decay (FID). Typical halftimes for the decay are .1 to 2 seconds for nuclear spins used for QIP. A normal NMR observation consists of measuring $M(t)$ at discrete time intervals until the signal is too small. The acquired FID is then Fourier-transformed to visualize the amplitudes of the different frequency contributions. The shape of the peaks in Figure 7 reflects the decay envelope. The width of the peaks is proportional to the decay rate λ .

For QIP, we wish to measure the probability p that a given qubit (say, qubit 1) is in the state $|1\rangle_1$. We have $1 - 2p = \text{tr}(\rho\sigma_z^{(1)})$, which is the expectation of $\sigma_z^{(1)}$. We can measure this expectation by first applying a 90° y -pulse to qubit 1 and thus changing the state to ρ' . This pulse has the effect of rotating initial, unobservable z -magnetization to observable x -magnetization. From $M(t)$ one can then infer $\text{tr}(\rho'\sigma_x^{(1)})$, which is the desired number. For the coupled pair of carbons, $\text{tr}(\rho'\sigma_x^{(1)})$ is given by the sum of the real components of the amplitudes of the 50 hertz and the -50 hertz contributions to $M(t)$. However, the problem is that these amplitudes are determined only up to a scale. A second problem is that the available states ρ are highly mixed (close to $\mathbb{1}/N$). The next section discusses how to compensate for both problems.

As a final comment on NMR measurement, note that the back reaction on the nuclear spins due to the emission of electromagnetic energy is weak. This is what enables us to measure the bulk magnetization over some time. The ensemble nature of the system gives direct, if noisy, access to expectations of observables such as σ_z rather than a single answer—0 or 1. For algorithms that provide a definite answer, having access only to expectations is not a problem because it is easy to distinguish the answer from the noise. However, using expectations can increase the need for quantum resources. For example, Shor’s factoring algorithm includes a significant amount of classical postprocessing based

on highly random answers from projective measurements. In order to implement the algorithm in an ensemble setting, the postprocessing must be performed reversibly and integrated into the quantum computation to guarantee a definite answer. Postprocessing can be done with polynomial additional quantum resources.

The Initial State. Because the energy difference between the nuclear spins' up and down states is so small compared with room temperature, the equilibrium distribution of states is nearly random. In the liquid samples used, equilibrium is established after 10 to 40 seconds if no rf fields are being applied. As a result, all computations start with the sample in equilibrium. One way to think of this initial state is that every nuclear spin in each molecule begins in the highly mixed state $(1 - \varepsilon)\mathbb{1}/2 + \varepsilon|\mathcal{O}\rangle\langle\mathcal{O}|$, where ε is a small number (of the order of 10^{-5}). This is a nearly random state with a small excess of the state $|\mathcal{O}\rangle$. The expression for the initial state derives from the fact that the equilibrium state ρ_{thermal} is proportional to $e^{-H/kT}$, where H is the internal Hamiltonian of the nuclear spins in a molecule (in energy units), T is the temperature, and k is the Boltzmann constant. In our case, H/kT is very small, and the coupling terms are negligible. Therefore,

$$e^{-H/kT} \approx e^{-\varepsilon_1\sigma_z^{(1)}/kT} e^{-\varepsilon_2\sigma_z^{(2)}/kT} \dots, \quad (13)$$

$$e^{-\varepsilon_1\sigma_z^{(1)}/kT} \approx \mathbb{1} - \varepsilon_1\sigma_z^{(1)}/kT, \text{ and} \quad (14)$$

$$e^{-H/kT} \approx \mathbb{1} - \varepsilon_1\sigma_z^{(1)}/kT - \varepsilon_2\sigma_z^{(2)}/kT - \dots, \quad (15)$$

where ε_l is half of the energy difference between the up and down states of the l^{th} nuclear spin.

Clearly, the available initial state is very far from what is needed for standard QIP. However, it can still be used to perform interesting computations. The main technique is to use available NMR tools to change the initial state to a pseudopure state, which for all practical purposes, behaves like the initial state required by QIP. The technique is based on three key observations. First, only the traceless part of the density matrix contributes to the magnetization. Suppose that we are using n spin-1/2 nuclei in a molecule and the density matrix is ρ . Then, the current magnetization is proportional to $\text{tr}(\rho \hat{m})$, where \hat{m} is a traceless operator—see Equation (9). Therefore, the magnetization does not depend on the part of ρ proportional to the identity matrix. A deviation density matrix for ρ is any matrix δ such that $\delta - \rho = \lambda \mathbb{1}$ for some λ . For example, $\varepsilon|\mathcal{O}\rangle\langle\mathcal{O}|$ is a deviation for the equilibrium state of one nuclear spin. We have

$$\begin{aligned} \text{tr}(\delta \hat{m}) &= \text{tr}((\rho + \lambda \mathbb{1}) \hat{m}) \\ &= \text{tr}(\rho \hat{m}) + \text{tr}(\hat{m}) \\ &= \text{tr}(\rho \hat{m}). \end{aligned} \quad (16)$$

The second observation is that all the unitary operations used, as well as the nonunitary ones to be discussed below, preserve the completely mixed state $\mathbb{1}/2^n$.¹ Therefore, all future observations of magnetization depend only on the initial deviation.

The third observation is that all the scales are relative. In particular, as will be explained, the probability that the final answer of a quantum computation is 1 can be

¹ The intrinsic relaxation process does not preserve the completely mixed state. But its contribution is either negligible over the time scale of typical experiments or can be removed with the help of subtractive phase cycling.

expressed as the ratio of two magnetizations. It follows that one can arbitrarily rescale a deviation density matrix. For measurement, the absolute size of the magnetizations is not important; the most important issue is that the magnetizations are strong enough to be observable over the noise.

To explain the relativity of the scales and introduce pseudopure states for QIP, we begin with one spin-1/2 qubit. Its equilibrium state has a deviation $\delta = \epsilon|0\rangle\langle 0|$. If U is the total unitary operator associated with a computation, then δ is transformed to $\delta = \epsilon U|0\rangle\langle 0|U^\dagger$. For QIP purposes, the goal is to determine what the final probability p_1 of measuring $|1\rangle$ is, given that $|0\rangle$ is the initial state. This probability can be computed as follows:

$$p_1 = \langle 1|U|0\rangle\langle 0|U^\dagger|1\rangle$$

$$= \text{tr}\left(U|0\rangle\langle 0|U^\dagger|1\rangle\langle 1|\right) \tag{17}$$

$$= \text{tr}\left(U|0\rangle\langle 0|U^\dagger(\mathbb{1} - \sigma_z)\right)/2 \tag{18}$$

$$= \left(\text{tr}\left(U|0\rangle\langle 0|U^\dagger\right) - \text{tr}\left(U|0\rangle\langle 0|U^\dagger\sigma_z\right)\right)/2 \tag{19}$$

$$= \left(1 - \text{tr}\left(U|0\rangle\langle 0|U^\dagger\sigma_z\right)\right)/2 . \tag{20}$$

Thus, the probability can be determined from the expectations of σ_z being measured for the initial and final states (in different experiments). This measurement yields the quantities $a = \text{tr}(\delta\sigma_z) = \epsilon$ and $a' = \text{tr}(\delta'\sigma_z) = \epsilon \text{tr}(U|0\rangle\langle 0|U^\dagger\sigma_z)$, respectively. The desired answer is $p_1 = (1 - (a/a'))/2$ and does not depend on the scale ϵ .

The method presented in the previous paragraph for determining the probability that the answer of a quantum computation is 1 generalizes to many qubits. The goal is to determine the probability p_1 of measuring $|1\rangle_1$ in a measurement of the first qubit after a computation with initial state $|0\dots 0\rangle$. Suppose we can prepare the spins in an initial state with a deviation $\delta = \epsilon|0\dots 0\rangle\langle 0\dots 0|$. A measurement of the expectations a and a' of σ_z^1 for the initial and final states then yields p_1 , as before, by the formula $p_1 = (1 - (a/a'))/2$.

A state with deviation $\epsilon|\psi\rangle\langle\psi|$ is called a pseudopure state because that deviation is proportional to the deviation of the pure state $|\psi\rangle\langle\psi|$. With respect to scale-independent NMR observations and unitary evolution, a pseudopure state is equivalent to the corresponding pure state. Because NMR QIP methods are scale independent, we now generalize the definition of deviation density matrix: δ is a deviation of the density matrix ρ if $\epsilon\delta = \rho + \lambda\mathbb{1}$ for some λ and ϵ .

Among the most important enabling techniques in NMR QIP are the methods that can be used to transform the initial thermal equilibrium state to a standard pseudopure state with deviation $|0\dots 0\rangle\langle 0\dots 0|$. An example of how that can be done will be given as the second algorithm in the section “Examples of Quantum Algorithms for NMR.” The basic principle for each method is to create, directly or indirectly by summing over multiple experiments, a new initial state as a sum $\rho_0 = \sum_i U_i \rho_{\text{thermal}} U_i^\dagger$, where the U_i are carefully and sometimes randomly chosen (Cory et al. 1997, Gershenfeld and Chuang 1997, Knill et al. 1998, Sharf et al. 2000) to ensure that ρ_0 has a standard pseudopure deviation. Among the most useful tools for realizing such sums are pulsed gradient fields.

Gradient Fields. Modern NMR spectrometers are equipped with the capability of applying a magnetic field gradient in any direction for a chosen, brief amount of time. If the direction is along the sample’s z -axis, then while the gradient is on, the field varies as $B(z) = B_0 + \gamma z B_1$, where B_0 is the strong, external field and B_1 is the gradient power.

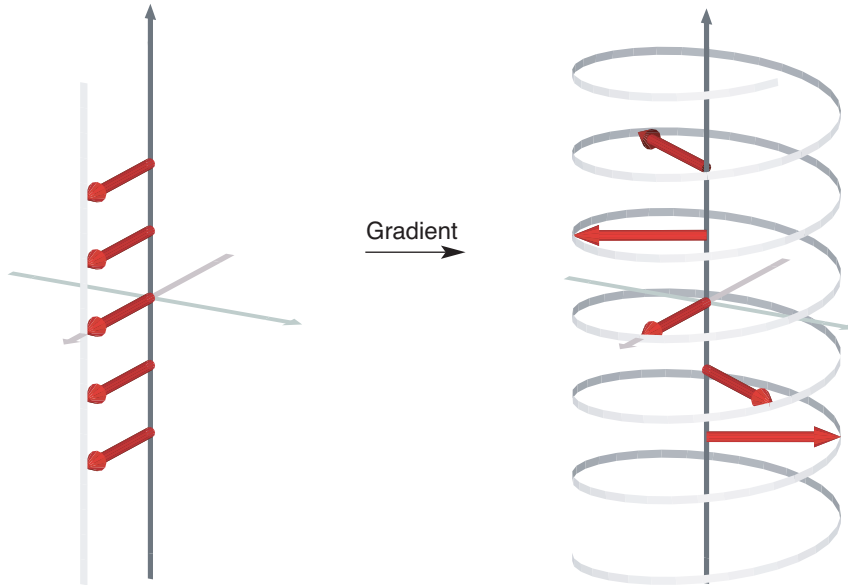


Figure 8. Pulsed Gradient Field along the z-Axis
Initial x -magnetization is assumed. A spin at $z = 0$ is not affected, but the ones above and below are rotated by an amount proportional to z . As a result, the local planar magnetization follows a spiral curve.

As a result of this gradient, the precession frequency of nuclear spins depends on their positions' z -coordinates. One of the most important applications of gradients is NMR imaging because gradients make it possible to distinguish different parts of the sample.

The effect of applying a z -gradient can be visualized for the situation in which there is only one observable nuclear spin per molecule. Suppose that the initial deviation density matrix of each nuclear spin is σ_x in the rotating frame. After a gradient pulse of duration t , the deviation of a nuclear spin at position z is given by $e^{-i\sigma_z \nu z t / 2} \sigma_x e^{i\sigma_z \nu z t / 2} = \cos(\nu z t) \sigma_x + \sin(\nu z t) \sigma_y$, where the constant ν depends linearly on the strength of the gradient and the magnetic moment of the nucleus—see Figure 8. The effect of the gradient is a z -dependent change in phase. The coil used to measure planar magnetization integrates the contribution to the magnetization of all the nuclei in the neighborhood of the coil. Assuming a coil equally sensitive over the interval between $-a$ and a along the sample's z -axis, the observed total x -magnetization is

$$\begin{aligned}
 M_x &= \int_{-a}^a dz \operatorname{tr} \left(\sigma_x (\cos(\nu z t) \sigma_x + \sin(\nu z t) \sigma_y) \right) \\
 &= \int_{-a}^a dz \operatorname{tr} \left(\cos(\nu z t) \sigma_x^2 + \sin(\nu z t) \sigma_x \sigma_y \right) \\
 &= \int_{-a}^a dz \operatorname{tr} \left(\cos(\nu z t) + i \sin(\nu z t) \sigma_z \right) \\
 &= 2 \int_{-a}^a dz \cos(\nu z t) .
 \end{aligned} \tag{21}$$

For large values of νt , $M_x \cong 0$. In general, a sufficiently powerful gradient pulse eliminates the planar magnetization.

Interestingly, the effect of a gradient pulse can be reversed if an opposite gradient pulse is applied for the same amount of time. This effect is called a “gradient echo.” The reversal only works if the second pulse is applied sufficiently soon. Otherwise, diffusion randomizes the molecules' positions along the gradient's direction before the second pulse. If the positions are randomized, the phase change from the second pulse is no longer correlated with that from the first for any given molecule. The loss of memory of the phase change from a gradient pulse can be fine-tuned by variations in the delay

between the two pulses in a gradient echo sequence. This method can be used for applying a controllable amount of phase noise, which is useful for investigating the effects of noise and the ability to correct for noise in QIP.

If the gradient pulse is not reversed and the memory of the phase changes is lost, then the pulse's effect can be described as an irreversible operation on the state of the nuclear spin. If the initial state of the nuclear spin in each molecule is ρ , then after the gradient pulse, the spin state of a molecule at position z is given by $\rho(z) = e^{-i\sigma_z vz t/2} \rho e^{i\sigma_z vz t/2}$. Suppose that the positions of the molecules are randomized over the region that the coil is sensitive to. Now it is no longer possible to tell where a given molecule was when the gradient pulse was applied. As a result, as far as our observations are concerned, the state of a molecule is given by $\rho(z)$, where z is random. In other words, the state is indistinguishable from

$$\rho' = \frac{1}{2a} \int_{-a}^a dz \rho(z) = \frac{1}{2a} \int_{-a}^a dz e^{-i\sigma_z vz t/2} \rho e^{i\sigma_z vz t/2} . \quad (22)$$

Thus, the effect of the gradient pulse is equivalent to the operation $\rho \rightarrow \rho'$ as defined by the above equation. This is an operation of the type mentioned at the end of the previous section and can be used for making states such as pseudopure states. Note that, after the gradients have been turned off, nuclei at different positions cannot be distinguished by the measurement coil. It is therefore not necessary to wait for the molecules' positions to be randomized.

So far, we have described the effects of gradient pulses on isolated nuclear spins in a molecule. In order to restrict the effect to a single nuclear spin in a molecule, one can invert the other spins between a pair of identical gradient pulses in the same direction. This technique refocuses the gradient for the inverted spins. An example of how effects involving multiple nuclear spins can be exploited is the algorithm for pseudopure state preparation described in the section "Creating a Labeled Pseudopure State."

Examples of Quantum Algorithms for NMR

We give three examples of algorithms for NMR QIP. The first is an NMR implementation of the controlled-not gate. The second consists of a procedure for preparing a type of pseudopure state. And the last shows how NMR can be used to investigate the behavior of simple error-correction procedures. The first two examples are fundamental to QIP with NMR. Realizations of the controlled-not are needed to translate standard quantum algorithms into the language of NMR, and procedures for making pseudopure states have to precede the implementation of many quantum algorithms.

The Controlled-not. One of the standard gates used in quantum algorithms is the controlled-not. The controlled-not gate (**cnot**) acts on two qubits. The action of **cnot** can be described by "if the first qubit is $|1\rangle$, then flip the second qubit." Consequently, the effect of **cnot** on the logical states is given by the mapping

$$\begin{aligned} \mathbf{cnot} |00\rangle &= |00\rangle \\ \mathbf{cnot} |01\rangle &= |01\rangle \\ \mathbf{cnot} |10\rangle &= |11\rangle \\ \mathbf{cnot} |11\rangle &= |10\rangle . \end{aligned} \quad (23)$$

As an operator, the controlled-not is given by

$$\mathbf{cnot} = |\mathbf{o}\rangle_1 \langle \mathbf{o}| + |\mathbf{1}\rangle_1 \langle \mathbf{1}| \sigma_x^{(2)} = \left((\mathbb{1} + \sigma_z^{(1)}) + (\mathbb{1} - \sigma_z^{(1)}) \sigma_x^{(2)} \right) / 2 . \quad (24)$$

The goal is to derive a sequence of NMR operations that realize the controlled-not. As discussed earlier (“Principles of Liquid-State NMR QIP”), the unitary operations implementable by simple NMR techniques are rotations $e^{-i\sigma_u^{(a)}\theta/2}$ by θ around the u -axis, where u is any direction in the plane (rf pulses), and the two-qubit operations $e^{-i\sigma_z^{(b)}\sigma_z^{(c)}\phi/2}$ (the J -coupling). We call $e^{-i\sigma_z^{(b)}\sigma_z^{(c)}\phi/2}$ a rotation by ϕ around $\sigma_z^{(b)}\sigma_z^{(c)}$. This terminology reflects the fact that such rotations and their effects on deviation density matrices can be understood by a generalization of the Bloch sphere picture called the product operator formalism introduced by Sørensen et al. (1983).

To implement the controlled-not using NMR techniques, one can decompose the gate into a sequence of 90° rotations around the main axes on each of the two qubits, and a 90° rotation around $\sigma_z^{(1)}\sigma_z^{(2)}$. One way to find a decomposition is to first realize that the two-qubit 90° rotation $e^{-i\sigma_z^{(1)}\sigma_z^{(2)}\pi/4}$ is equivalent to a combination of two gates, each conditional on the logical state of qubit 1. The first gate applies a 90° rotation around the z -axis ($e^{-i\sigma_z^{(2)}\pi/4}$) to qubit 2 conditional on qubit 1’s state being $|\mathbf{o}\rangle_1$. The second applies the -90° rotation $e^{i\sigma_z^{(2)}\pi/4}$ to qubit 2 conditional on qubit 1’s state being $|\mathbf{1}\rangle_1$. By following the two-qubit rotation with a -90° rotation around the z -axis ($e^{i\sigma_z^{(2)}\pi/4}$) on qubit 2, the total effect is to cancel the rotation if qubit 1 is in state $|\mathbf{o}\rangle_1$; if qubit 1 is in state $|\mathbf{1}\rangle_1$, the rotations add to a -180° rotation $e^{i\sigma_z^{(2)}\pi/2} = i\sigma_z^{(2)}$ on qubit 2. If we precede this sequence with $e^{-i\sigma_y^{(2)}\pi/4}$ and follow it by $e^{i\sigma_y^{(2)}\pi/4}$ (this operation is called conjugating by a -90° y -rotation), the overall effect is a conditional $-i\sigma_x^{(2)}$ operation. Note how the conjugation rotated the operation’s axis according to the Bloch sphere rules. The controlled-not is obtained by eliminating the $-i$ with a 90° z -rotation on qubit 1. That is, the effect of the complete sequence is $e^{-i\pi/4}|\mathbf{o}\rangle_1 \langle \mathbf{o}| + e^{-i\pi/4}|\mathbf{1}\rangle_2 \langle \mathbf{1}| \sigma_x^{(2)}$, which is the controlled-not up to a global phase. The decomposition thus obtained can be represented as a quantum network with rotation gates, as shown in Figure 9. The corresponding NMR pulse sequence implementation is shown in Figure 10.

The effect of the NMR pulse sequence that implements the controlled-not can be visualized for logical initial states with the help of the Bloch-sphere representation of the states. Figure 11 shows such a visualization for two initial states.

The effects of the pulse sequence for the controlled-not can be shown with the Bloch sphere (Figure 11) only if the intermediate states are products of states on each qubit. Things are no longer so simple if the initial state of the spins is $1/\sqrt{2}(|\mathbf{o}\rangle + |\mathbf{1}\rangle) |\mathbf{o}\rangle = 1/\sqrt{2}(|\mathbf{oo}\rangle + |\mathbf{1o}\rangle)$, for example. This is representable as spin 1’s arrow pointing along the x -axis, but the J -coupling leads to a superposition of states (a maximally entangled state) no longer representable by a simple combination of arrows in the Bloch sphere.

Creating a Labeled Pseudopure State. One way to realize the standard pseudopure state starting from the equilibrium density matrix ρ_{thermal} is to eliminate the observable contributions due to terms of ρ_{thermal} different from $|\mathbf{o}\dots\mathbf{o}\rangle\langle\mathbf{o}\dots\mathbf{o}|$. There are several different methods of accomplishing this task. For example, one can perform multiple experiments with different preprocessing of the equilibrium state so that signals from unwanted terms average to zero (temporal averaging), or one can use gradients to remove the unwanted terms in one experiment (spatial averaging).

In this section, we show how to use spatial averaging to prepare a so-called labeled pseudopure state on two nuclear spins. In general, instead of preparing the standard pseudopure state with deviation $|\mathbf{o}\dots\rangle\langle\mathbf{o}\dots|$ on n spin-1/2 nuclei, one can prepare a

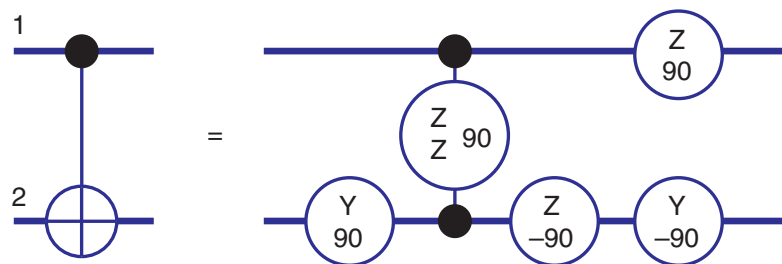


Figure 9. Quantum Network for Implementing the Controlled-not with NMR Operations

The conventions for depicting gates are as explained in the article “Quantum Information Processing” on page 2. The two one-qubit z-rotations can be implemented by a change in the reference phase of the rotating frame without any rf pulses being applied.

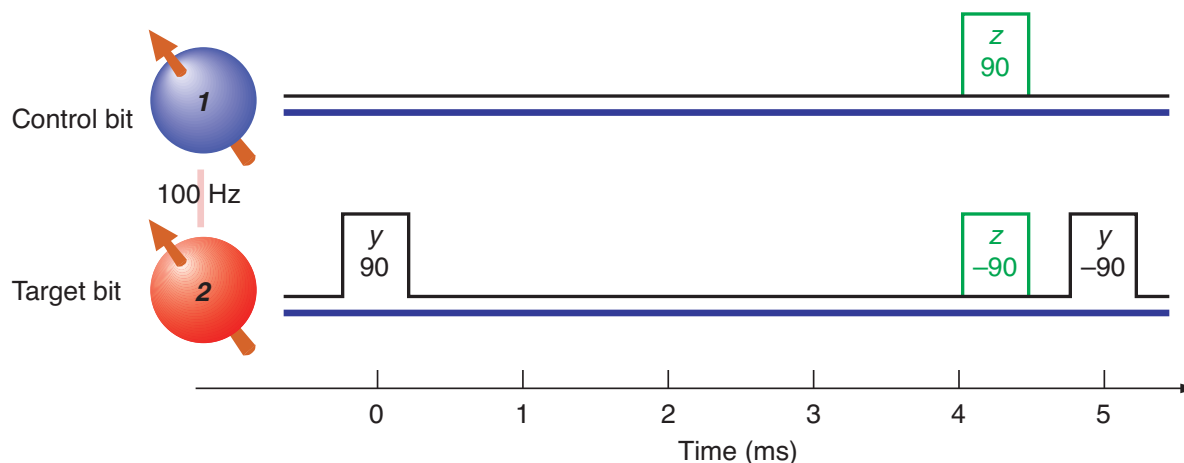


Figure 10. Pulse Sequence for Realizing the Controlled-not

The control bit is spin 1 and the target is spin 2. The pulses are shown with the representation introduced in Figure 6. The z-pulses (shown in green) are virtual, requiring only a change of reference frame. The placement of the z-pulses between the rf pulses is immaterial because they commute with the coupling that evolves in between. The delay between the two rf pulses is $1/(2J)$ (5 ms if $J = 100$ Hz), which realizes the desired two-qubit rotation by internal evolution. The -90° y-rotation is actually implemented with a 90° pulse with axis $-y$. The resulting rotation has the desired effect up to a global phase. The pulse widths are exaggerated and should be as short as possible to avoid errors due to coupling evolution during the rf pulses. Alternatively, techniques can be used that compensate for some of these errors (Knill et al. 2000).

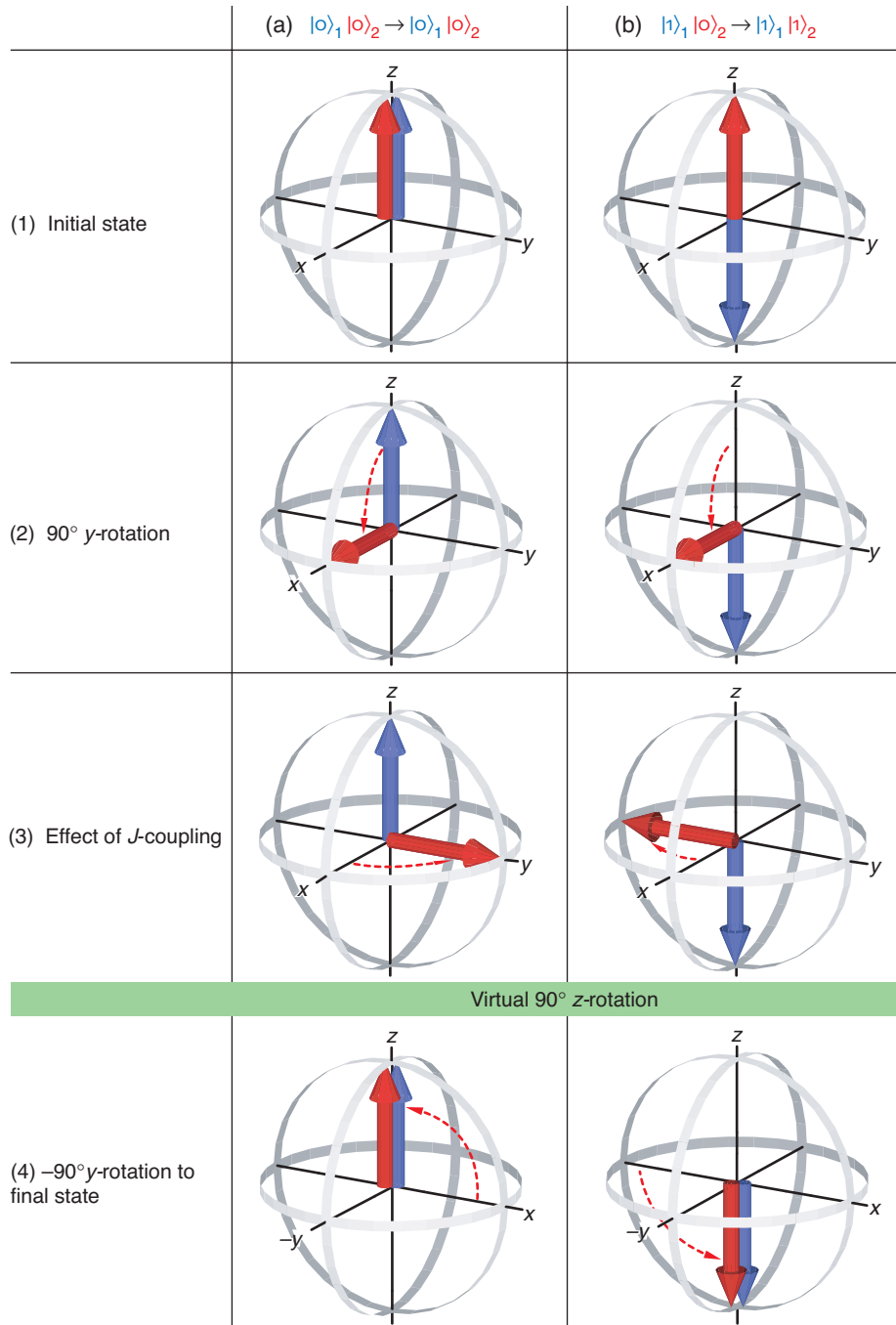


Figure 11. States Corresponding to the Controlled-not Pulse Sequence

The two columns (a) and (b) show the evolution of the qubit states during the controlled-not pulse sequence. The blue and red arrows represent spin 1 and 2, respectively. The configurations in rows 1 to 4 are shown (1) at the beginning of the sequence, (2) after the 90° y-rotation, (3) after the *J*-coupling (but before the z- and y-pulses), and (4) at the end of the sequence. The conditional effect is realized by the second spin's pointing down at the end of the second column. The effect of the *J*-coupling causing the evolution from 2 to 3 is best understood as a conditional rotation around the z-axis (forward by 90° if the first spin is up; backward, if it is down).

labeled pseudopure state with deviation $\sigma_x^{(1)}|0\dots\rangle\langle 0\dots|$ on $n + 1$ spins. This state is easily recognizable with an NMR observation of the first spin: Assuming that all the peaks arising from couplings to other spins are resolved, the first spin's peak group has 2^n peaks, corresponding to which logical states the other spins are in. If the current state is the labeled pseudopure state just mentioned, then all the other spins are in the logical state $|0\rangle$, which implies that, in the spectrum, only one of the peaks of the first spin's peak group is visible (see Figure 12).

The labeled pseudopure state can be used as a standard pseudopure state on n qubits. Observation of the final answer of a computation is possible by observing spin 1, provided that the coupling to the answer-containing spin is sufficiently strong for the peaks corresponding to its two logical states to be well separated. For this purpose, the couplings to the other spins need not be resolved in the peak group. Specifically, to determine the answer of a computation, the peaks of the spin 1 peak group are separated into two subgroups, the first (second) containing the peaks associated with the answer-containing spin being in state $|0\rangle$ ($|1\rangle$), respectively. Comparing the total signal in each of the two peak subgroups gives the relative probabilities of the two answers (0 or 1).

The labeled pseudopure state can also be used to investigate the effect of a process that manipulates the state of one qubit and requires n additional initialized qubits. Examples include experimental verification of one-qubit error-correcting codes as explained in the next section.

For preparing the two-qubit labeled pseudopure state, consider the two carbon nuclei in labeled TCE with the proton spin decoupled so that its effect can be ignored. A "transition" in the density matrix for this system is an element of the density matrix of the form $|ab\rangle\langle cd|$, where a, b, c , and d are 0 or 1 . Let $\Delta(ab, cd) = (a - c) + (b - d)$, where in the expression on the right, a, b, c , and d are interpreted as the numbers 0 or 1 , as appropriate. Applying a pulsed gradient along the z -axis evolves the transitions according to $|ab\rangle\langle cd| \rightarrow e^{i\Delta(ab,cd)vz}|ab\rangle\langle cd|$, where v is proportional to the product of the gradient power and pulse time and z is the molecule's position along the z -coordinate. For example, $|01\rangle\langle 10|$ has $\Delta = 0$ and is not affected whereas $|00\rangle\langle 11|$ acquires a phase of e^{-i2vz} . There are only two transitions, $|00\rangle\langle 11|$ and $|11\rangle\langle 00|$, whose acquired phase has a rate of $\Delta = \pm 2$ along the z -axis. These transitions are called 2-coherences because $\Delta = \pm 2$. The idea is to first recognize that these transitions can be used to define a labeled pseudopure "cat" state (see below), then to exploit the 2-coherences' unique behavior under the gradient in order to extract the pseudopure cat state, and finally to decode to a standard labeled pseudopure state. Note that the property that 2-coherences' phases evolve at twice the basic rate is a uniquely quantum phenomenon for two spins. No such effect is observed for a pair of classical spins.

The standard two-qubit labeled pseudopure state's deviation can be written as $\rho_{\text{std}_x} = \sigma_x^{(1)}1/2(1 + \sigma_z^{(2)})$. We can consider other deviations of this form where the two Pauli operators are replaced by a pair of different commuting products of Pauli operators. An example is

$$\rho_{\text{cat}_x} = \left(\sigma_x^{(1)}\sigma_x^{(2)}\right) \frac{1}{2} \left(1 + \sigma_z^{(1)}\sigma_z^{(2)}\right), \tag{25}$$

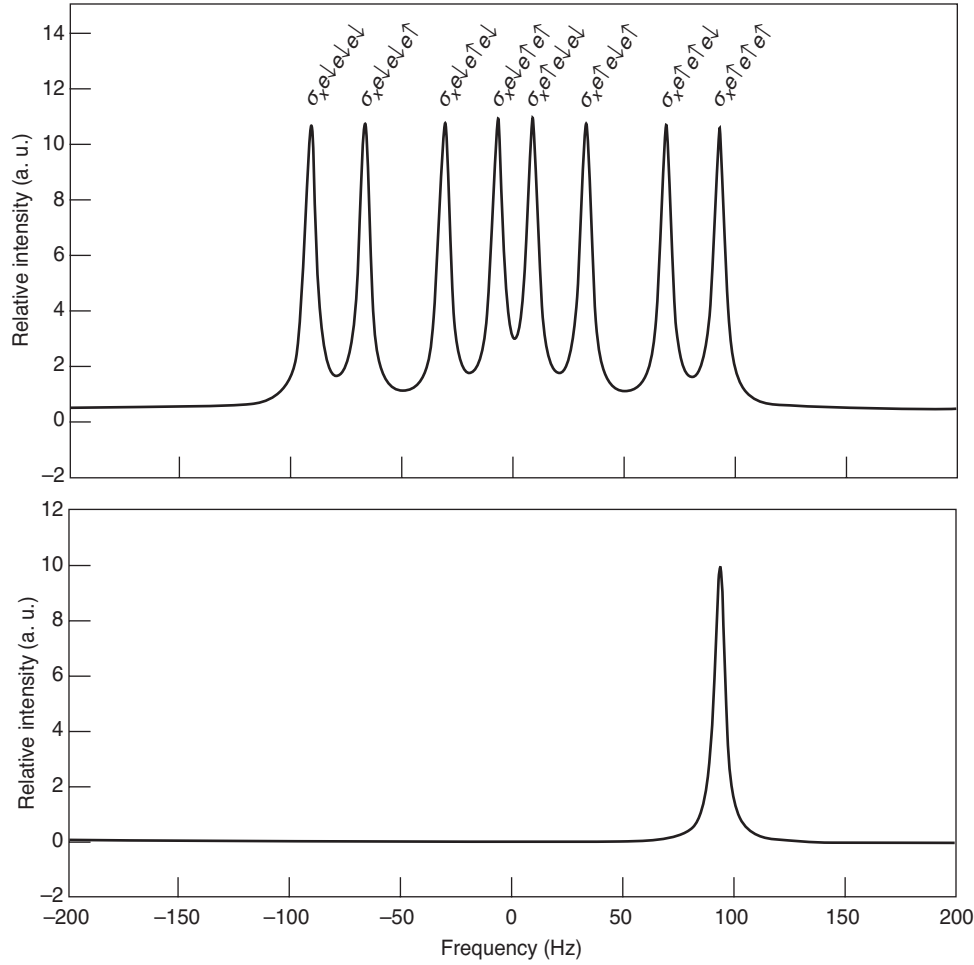


Figure 12. Labeled Pseudopure State Spectrum vs Peak Group
 (a) This spectrum shows the peak group of a simulated nuclear spin coupled to three other spins with coupling constants of 100 Hz, 60 Hz, and 24 Hz. The simulation parameters are the same as in Figure 7. Given above each peak is the part of the initial deviation that contributes to the peak. The spin labels have been omitted. Each contributing deviation consists of σ_x on the observed nucleus followed by one of the logical (up or down) states (density matrices) for each of the other spins. The notation is as defined after Equation 9.
 (b) This spectrum shows what is observed if the initial deviation is the standard labeled pseudopure state. This state contributes only to the rightmost peak, as this peak is associated with the logical $|o\rangle$ states on the spins not observed.

where we replaced $\sigma_x^{(1)}$ by $\sigma_x^{(1)}\sigma_x^{(2)}$ and $\sigma_z^{(2)}$ by $\sigma_z^{(1)}\sigma_z^{(2)}$. As announced, the two Pauli products commute. We will show that there is a simple sequence of 90° rotations whose effect is to decode the deviations $\sigma_x^{(1)}\sigma_x^{(2)} \rightarrow \sigma_x^{(1)}$ and $\sigma_z^{(1)}\sigma_z^{(2)} \rightarrow \sigma_z^{(2)}$, thus converting the state ρ_{cat_x} to ρ_{std_x} . The state ρ_{cat_x} can be expressed in terms of the transitions as follows:

$$\rho_{\text{cat}_x} = |oo\rangle\langle 11| + |11\rangle\langle oo|. \tag{26}$$

It can be seen that ρ_{cat_x} consists only of 2-coherences. Another such state is

$$\rho_{\text{cat}_y} = \left(\sigma_x^{(1)}\sigma_y^{(2)}\right) \frac{1}{2} \left(\mathbb{1} + \sigma_z^{(1)}\sigma_z^{(2)}\right) \tag{27}$$

$$= -i|oo\rangle\langle 11| + i|11\rangle\langle oo|. \tag{28}$$

Suppose that one can create a state that has a deviation of the form $\rho = \alpha\rho_{\text{cat}_x} + \beta\rho_{\text{rest}}$ such that ρ_{rest} contains no 2-coherences or 0-coherences. After a gradient pulse is applied, the state becomes

$$\alpha\left(\cos(2vz)\rho_{\text{cat}_x} + \sin(2vz)\rho_{\text{cat}_y}\right) + \beta\rho_{\text{rest}}(z) , \quad (29)$$

where $\rho_{\text{rest}}(z)$ depends periodically on z with spatial frequencies of $\pm v$, not $\pm 2v$ or 0. We can then decode this state to

$$\rho = \alpha\left(\cos(2vz)\rho_{\text{std}_x} + \sin(2vz)\rho_{\text{std}_y}\right) + \beta\rho'_{\text{rest}}(z) \quad (30)$$

$$= \alpha\left(\cos(2vz)\sigma_x^{(1)} + \sin(2vz)\sigma_y^{(1)}\right)\frac{1}{2}\left(\mathbb{1} + \sigma_z^{(1)}\right) + \beta\rho'_{\text{rest}}(z) . \quad (31)$$

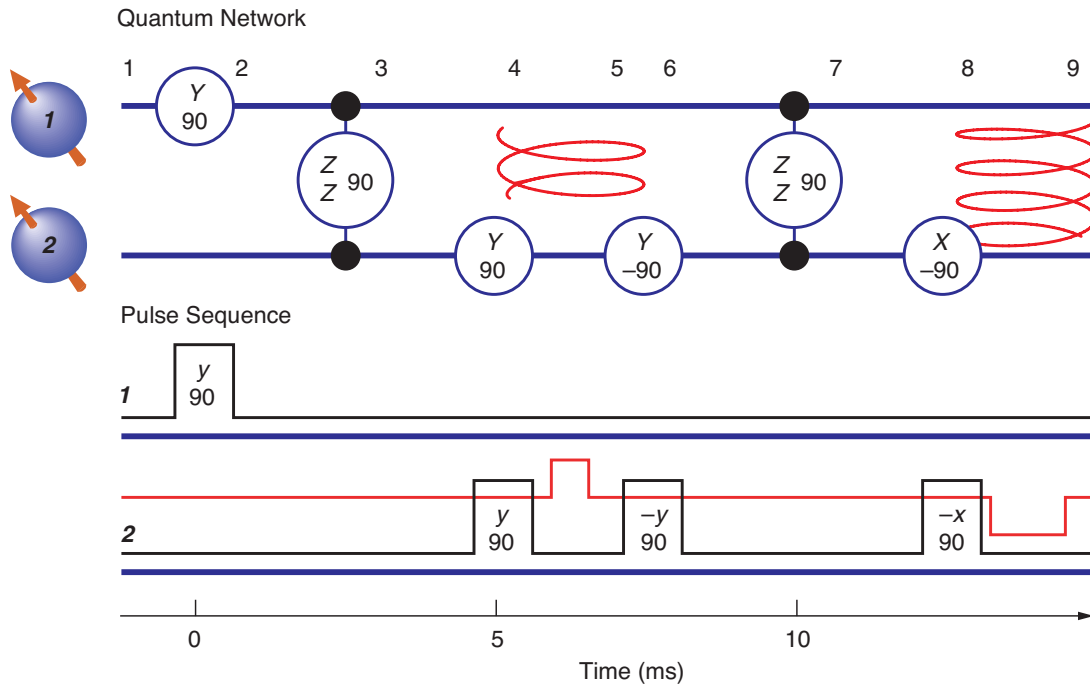
If one now applies a gradient pulse of twice the total strength and opposite orientation, the first term is restored to $\alpha\rho_{\text{std}_x}$, but the second term retains nonzero periodicities along z . Thus, if we no longer use any operations to distinguish among different molecules along the z -axis or if we let diffusion erase the memory of the position along z , then the second term is eliminated from observability by being averaged to 0. The desired labeled pseudopure state is obtained. Zero-coherences during the initial gradient pulse are acceptable provided that the decoding transfers them to coherences different from 0 or 2 during the final pulse in order to ensure that they also average to 0. A pulse sequence that realizes a version of the above procedure is shown in Figure 13.

We can follow what happens to an initial deviation density matrix of $\sigma_z^{(1)}$ as the network of Figure 13 is executed. We use product operators with the abbreviations $I = \mathbb{1}$, $X = \sigma_x$, $Y = \sigma_y$, $Z = \sigma_z$ and, for example, $XY = \sigma_x^{(1)}\sigma_y^{(2)}$. At the checkpoints indicated in the figure, the deviations are the following:

Checkpoints	1	ZI	
	2	XI	
	3	YZ	
	4	$YX \propto$	
		$YX + XY$	$+ YX - XY$
	5	$\cos(2vz)(YX + XY) + \sin(2vz)(YY - XX)$	$+ YX - XY$
	6	$\cos(2vz)(YZ + XY) + \sin(2vz)(YY - XZ)$	$+ YZ - XY$
	7	$\cos(2vz)(-XI + XY) + \sin(2vz)(YY - YI)$	$+ -XI - XY$
	8	$\cos(2vz)(-XI - XZ) + \sin(2vz)(-YZ - YI)$	$+ -XI + XZ$
	9	$-X(I + Z)$	$+ -(\cos(-2vz)X + \sin(-2vz)Y)(I - Z) .$

Except for a sign, the desired state is obtained. The rightmost term is eliminated after integrating over the sample or after diffusion erases memory of z .

This method for making a two-qubit labeled pseudopure state can be extended to arbitrarily many (n) qubits by exploiting the two n -coherences, which are the transitions with $\Delta = \pm n$. An experiment implementing this method can be used to determine how good the available quantum control is. The quality of the control is determined by a comparison of two spectral signals: I_p , the intensity of the single peak that shows up in the peak group for spin 1 when observing the labeled pseudopure state, and I_0 , the intensity of the same peak in an observation of the initial deviation after applying a 90° pulse



to rotate $\sigma_z^{(1)}$ into the plane. We performed this experiment on a seven-spin system and determined that $I_p/I_0 = .73 \pm .02$. This result implies a total error of 27 ± 2 percent. Because the implementation has 12 two-qubit gates, an error rate of about 2 percent per two-qubit gate is achievable for nuclear spins in this setting (Knill et al. 2000).

Quantum Error Correction for Phase Errors. Currently envisaged scalable quantum computers require the use of quantum error correction to enable relatively error-free computation on a platform of physical systems that are inherently error prone. For this reason, some of the most commonly used subroutines in quantum computers will be associated with maintaining information in encoded forms. This observation motivates experimental realizations of quantum error correction to determine whether adequate control can be achieved in order to implement these subroutines and to see in a practical setting that error correction has the desired effects. Experiments to date have included realizations of a version of the three-qubit repetition code (Cory et al. 1998) and of the five-qubit one-error-correcting code (the shortest possible such code)—see the article “Quantum Information Processing” on page 2. In this section, we discuss the experimental implementation of the former.

In NMR, one of the primary sources of error is phase decoherence of the nuclear spins due to both systematic and random fluctuations in the field along the z -axis. At the same time, using gradient pulses and diffusion, phase decoherence is readily induced artificially and in a controlled way. The three-bit quantum repetition code (see the article “Introduction to Quantum Error Correction” on page 188) can be adapted to protect against phase errors to first order. Define $|+\rangle = 1/\sqrt{2} (|0\rangle + |1\rangle)$ and $|-\rangle = 1/\sqrt{2} (|0\rangle - |1\rangle)$. The code we want is defined by the logical states

$$|0\rangle_L = |+\rangle|+\rangle|+\rangle, \text{ and } |1\rangle_L = |-\rangle|-\rangle|-\rangle. \tag{33}$$

It is readily seen that the three one-qubit phase errors $\sigma_z^{(1)}$, $\sigma_z^{(2)}$, and $\sigma_z^{(3)}$ and “no error” (1) unitarily map the code to orthogonal subspaces. It follows that this set of

Figure 13. Realizing a Two-Qubit Labeled Pseudopure State
 The network is shown above the pulse sequence realizing it. A coupling constant of 100 Hz is assumed. Gradients are indicated by spirals in the network. The gradient strength is given as the red line in the pulse sequence. The doubling of the integrated gradient strength required to achieve the desired “echo” is indicated by a doubling of the gradient pulse time. The numbers above the quantum network are checkpoints used in the discussion below. The input state’s deviation is assumed to be $\sigma_z^{(1)}$. This deviation can be obtained from the equilibrium state by applying a 90° rotation to spin 2 followed by a gradient pulse along another axis to remove $\sigma_z^{(2)}$. Instead of using a gradient pulse, one can use phase cycling, which involves performing two experiments, the second having the sign of the phase in the first y -pulse changed, and then subtracting the measured signals.

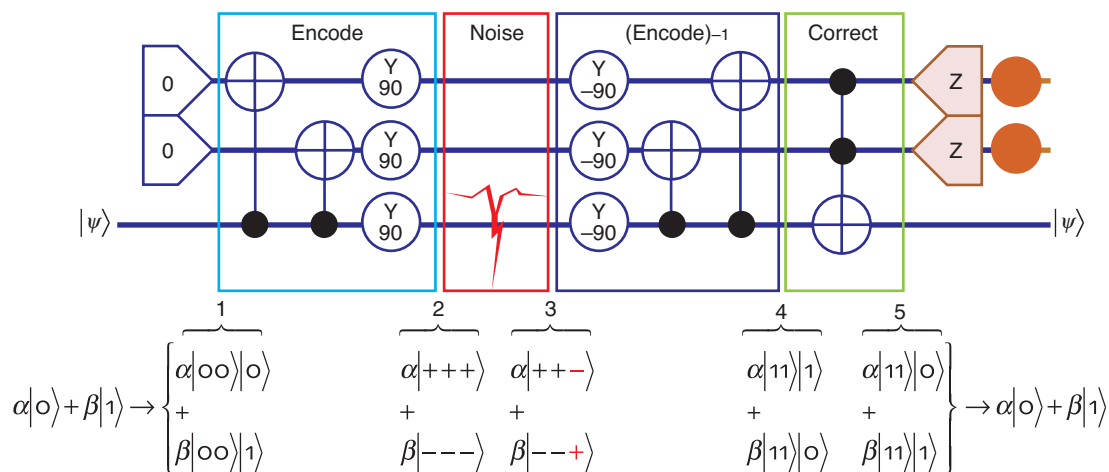


Figure 14. Quantum Network for the Three-Qubit Phase-Error-Correcting Repetition Code

The bottom qubit is encoded with two controlled-nots and three y-rotations. In the experiment, either physical or controlled noise is allowed to act. The encoded information is then decoded. For the present purposes, it is convenient to separate the decoding procedures into two steps: The first is the inverse of the encoding procedure; the second consists of a Toffoli gate that uses the error information in the syndrome qubits (the top two) to restore the encoded information. The Toffoli gate in the last step flips the output qubit conditionally on the syndrome qubits' state being $|11\rangle$. This gate can be realized with NMR pulses and delays by using more sophisticated versions of the implementation of the controlled-not. The syndrome qubits can be "dumped" at the end of the procedure. The behavior of the network is shown for a generic state in which the bottom qubit experiences a σ_z error.

errors is correctable (for a full discussion, see the article "Introduction to Error Correction" on page 188). The simplest way to use this code is to encode one qubit's state into it, wait for some errors to happen, and then decode to an output qubit. Success is indicated by the output qubit's state being significantly closer to the input qubit's state after error correction. Without errors between encoding and decoding, the output state should be the same as the input state, provided that the encoding and decoding procedures are implemented perfectly. Therefore, in this case, the experimentally determined difference between input and output gives a measurement of how well the procedures were implemented.

To obtain the phase-correcting repetition code from the standard repetition code, we apply Hadamard transforms or 90° y-rotations to each qubit. The quantum network shown in Figure 14 was obtained in this fashion from the network given in the article on error correction.

To determine the behavior and the quality of the implementation for various σ_z -error models in an actual NMR realization, one can use as initial states labeled pseudopure states with deviations $\sigma_u|00\rangle\langle 00|$ for $u = x, y, z$. Without error, the total output signal on spin 1 along σ_u for each u should be the same as the input signal. Some of the data reported by Cory and coworkers (1998) are shown in Figure 15.

Work on benchmarking error-control methods using liquid-state NMR is continuing. Other experiments include the implementation of a two-qubit code with an application to phase errors (Leung et al. 1999) and the verification of the shortest nontrivial noiseless subsystem on three qubits (Viola et al. 2001). The latter demonstrates that, for some physically realistic noise models, it is possible to store quantum information in such a way that it is completely unaffected by the noise.

Discussion

Overview of Contributions to QIP. Important issues in current experimental efforts toward realizing QIP are to find ways of achieving necessary quantum control and to determine whether sufficiently low error rates are possible. Liquid-state NMR is the only extant system (as of 2002) with the ability to realize relatively universal manipulations on more than two qubits—restricted control has been demonstrated in four ions (Sackett et al 2000). For this reason, NMR serves as a useful platform for developing and experimentally verifying techniques for QIP and for establishing simple procedures for benchmarking information-processing tasks. The cat state and the various error-

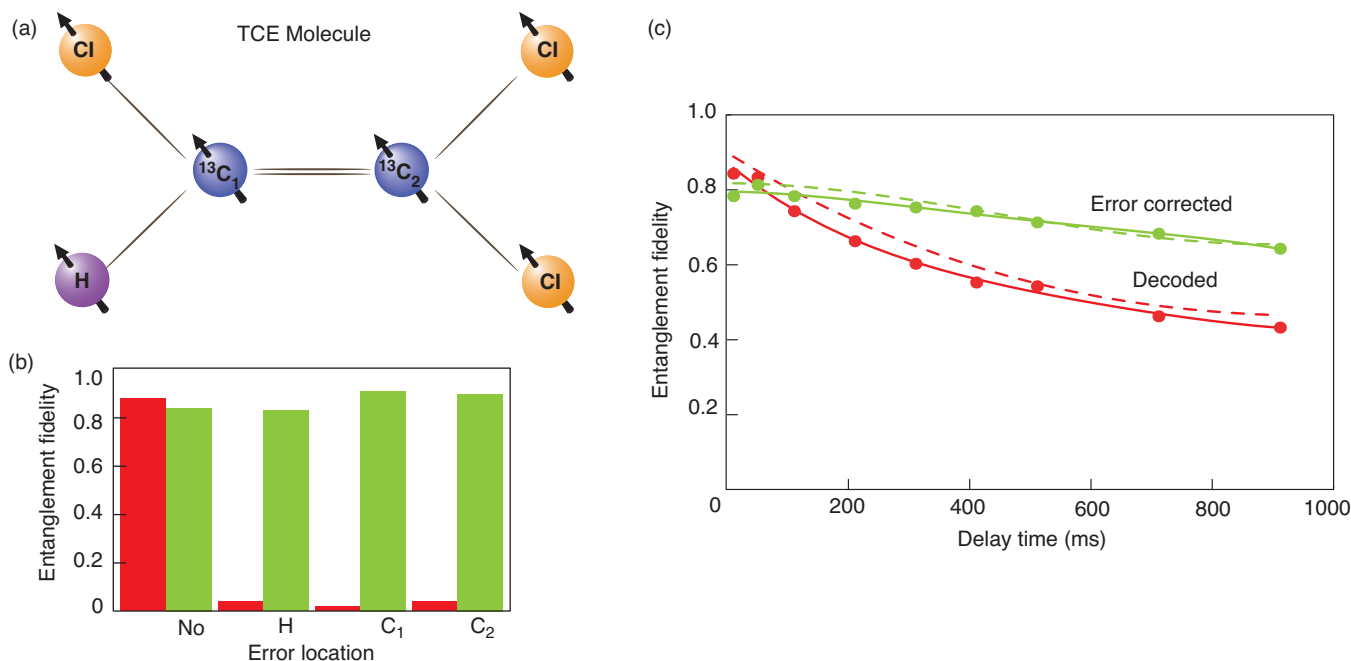


Figure 15. Experimental Fidelities of Error Correction

(a) The molecule used in the experiment is shown here. (b) The bar graph shows fidelities for explicitly applied errors. The fidelities f (technically, the entanglement fidelities) are an average of the signed ratios f_u of the input to the output signals for the initial deviations $\sigma_u|00\rangle\langle 00|$ with $u = x, y, z$. Specifically, $f = 1/4(1 + f_x + f_y + f_z)$. The reduction from 1 of the green bars (showing fidelity for the full procedure) is due to errors in our implementation of the pulses and from relaxation processes. The red bars are the fidelity for the output before the last error-correction step, and they contain the effects of the errors. (c) The graph shows the fidelities for the physical relaxation process. Here, the evolution consisted of a delay of up to 1000 ms. The red curve is the fidelity of the output qubit before the final Toffoli gate that

corrects the errors based on the syndrome. The green curve is the fidelity of the output after the Toffoli gate. The effect of error correction can be seen by a significant flattening of the curve because correction of first-order (that is, single) phase errors implies that residual, uncorrected (double or triple) phase errors increase quadratically in time. The green curve starts lower than the red one because of additional errors incurred by the implementation of the Toffoli gate. The dashed curves are obtained by simulation using estimated phase relaxation rates with halflives of 2 s (proton), 0.76 s (first carbon) and 0.42 s (second carbon). Errors in the data points are approximately 0.05. (For a more thorough implementation and analysis of a three-qubit phase-error-correcting code, see Sharf et al. 2000).

correction benchmarks (Knill et al. 2000, Knill et al. 2001) consist of a set of quantum control steps and measurement procedures that can be used with any general-purpose QIP system to determine, in a device-independent way, the degree of control achieved. The demonstration of error rates in the few percent per nontrivial operation is encouraging. For existing and proposed experimental systems other than NMR, achieving such error rates is still a great challenge.

Prior research in NMR, independent of quantum information, has proved to be a rich source of basic quantum-control techniques useful for physically realizing quantum information in other settings. We mention four examples. The first is the development of sophisticated shaped-pulse techniques that can selectively control transitions or spins while being robust against typical errors. These techniques are finding applications to quantum control involving laser pulses (Warren et al. 1993) and are likely to be very useful when using coherent light to accurately control transitions in atoms or quantum dots, for example. The second is the recognition that there are simple ways in which imperfect pulses can be combined to eliminate systematic errors such as those associated with miscalibration of power or side effects on off-resonant nuclear spins. Although

many of these techniques were originally developed for such problems as accurate inversion of spins, they are readily generalized to other quantum gates (Levitt 1982, Cummins and Jones 1999). The third example is decoupling used to reduce unwanted external interactions. For example, a common problem in NMR is to eliminate the interactions between proton and labeled carbon nuclear spins in order to observe decoupled carbon spins. In this case, the protons constitute an external system with an unwanted interaction. To eliminate the interaction, it is sufficient to invert the protons frequently. Sophisticated techniques for ensuring that the interactions are effectively turned off independent of pulse errors have been developed (Ernst et al. 1994). These techniques have been greatly generalized and shown to be useful for actively creating protected qubit subsystems in any situation in which the interaction has relatively long correlation times (Viola and Lloyd 1998, Viola et al. 1999). Refocusing to undo unwanted internal interactions is our fourth example. The technique for turning off the coupling between spins that is so important for realizing QIP in liquid-state NMR is a special case of much more general methods of turning off or refocusing Hamiltonians. For example, a famous technique in solid-state NMR is to reverse the dipolar coupling Hamiltonian using a clever sequence of 180° pulses at different phases (Ernst et al. 1994, page 48). Many other proposed QIP systems suffer from such internal interactions while having similar control opportunities.

The contributions of NMR QIP research extend beyond those directly applicable to experimental QIP systems. It is due to NMR that the idea of ensemble quantum computation with weak measurement was introduced and recognized as being, for true pure initial states, as powerful for solving algorithmic problems as the standard model of quantum computation. (It cannot be used in settings involving quantum communication.) One implication is that, to a large extent, the usual assumption of projective measurement can be replaced by any measurement that can statistically distinguish between the two states of a qubit. Scalability still requires the ability to reset qubits during the computation, which is not possible in liquid-state NMR. Another interesting concept emerging from NMR QIP is that of computational cooling (Schulman and Vazirani 1998), which can be used to efficiently extract initialized qubits from a large number of noisy qubits in initial states that are only partially biased toward $|0\rangle$. This is a very useful tool for better exploiting otherwise noisy physical systems.

The last example of interesting ideas arising from NMR studies is the one-qubit model of quantum computation (Knill and Laflamme 1998). This is a useful abstraction of the capabilities of liquid-state NMR. In this model, it is assumed that initially, one qubit is in the state $|0\rangle$ and all the others are in random states. Standard unitary quantum gates can be applied, and the final measurement is destructive. Without loss of generality, one can assume that all qubits are reinitialized after the measurement. This model can perform interesting physics simulations with no known efficient classical algorithms. On the other hand, with respect to oracles, it is strictly weaker than quantum computation. It is also known that it cannot faithfully simulate quantum computers (Ambainis et al. 2000).

Capabilities of Liquid-State NMR. One of the main issues in liquid-state NMR QIP is the highly mixed initial state. The methods for extracting pseudopure states are not practical for more than 10 (or so) nuclear spins. The problem is that for these methods, the pseudopure state signal decreases exponentially with the number of qubits prepared while the noise level is constant. This exponential loss limits the ability to explore and benchmark standard quantum algorithms even in the absence of noise. There are in fact ways in which liquid-state NMR can be usefully applied to many more qubits. The first and less practical is to use computational cooling for a (unrealistically) large number of spins to obtain less mixed initial states. Versions of this technique have been studied and used in NMR to increase signal to noise (Glaser et al. 1998). The second is to use the

one-qubit model of quantum computation instead of trying to realize pseudopure states. For this purpose, liquid-state NMR is limited only by relaxation noise and pulse control errors, not by the number of qubits. Noise still limits the number of useful operations, but nontrivial physics simulations are believed to be possible with less than 100 qubits (Lloyd 1996). Remarkably, a one-qubit quantum computer can efficiently obtain a significant amount of information about the spectrum of a Hamiltonian that can be emulated on a quantum computer (Knill and Laflamme 1998, Somma et al. 2002, Miquel et al. 2002). Consequently, although QIP with molecules in liquid state cannot realistically be used to implement standard quantum algorithms involving more than about 10 qubits, its capabilities have the potential of exceeding the resource limitations of available classical computers for some applications.

Prospects for NMR QIP. There are many more algorithms and benchmarks that can be usefully explored using the liquid state NMR platform. We hope to soon have a molecule with ten or more useful spins and good properties for QIP. Initially, this molecule can be used to extend and verify the behavior of existing scalable benchmarks. Later, experiments testing basic ideas in physics simulation or more sophisticated noise-control methods are likely.

Liquid-state NMR QIP is one of many ways in which NMR can be used for quantum information. One of the promising proposals for quantum computation is based on phosphorus embedded in silicon (Kane 1998) and involves controlling phosphorus nuclear spins using NMR methods. In this proposal, couplings and frequencies are controlled with locally applied voltages. Universal control can be implemented with rf pulses. It is also possible to scale up NMR QIP without leaving the basic paradigms of liquid-state NMR while adding such features as high polarization, the ability to dynamically reset qubits (required for scalability), and much faster two-qubit gates. One proposal for achieving this goal is to use dilute molecules in a solid-state matrix instead of molecules in liquid (Cory et al. 2000). This approach may lead to pure-state quantum computation for significantly more than ten qubits.

NMR QIP has been a useful tool for furthering our understanding of the experimental challenges of quantum computation. We believe that NMR QIP will continue to shed light on important issues in physically realizing quantum information. ■

Further Reading

- Ambainis, A., L. J. Schulman, and U. Vazirani. 2000. Computing with Highly Mixed States. In *Proceedings of the 32th Annual ACM Symposium on the Theory of Computation (STOC)*. p. 697. New York: ACM Press.
- Anderson, A. G., and E. L. Hahn. 1955. Spin Echo Storage Technique. U.S. Patent # 2,714,714.
- Anderson, A. G., R. Garwin, E. L. Hahn, J. W. Horton, and G. L. Tucker. 1955. Spin Echo Serial Storage Memory. *J. Appl. Phys.* **26**: 1324.
- Barenco, A., C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, et al. 1995. Elementary Gates for Quantum Computation. *Phys. Rev. A* **52**: 3457.
- Bloch, F. 1946. Nuclear Induction. *Phys. Rev.* **70**: 460.
- Chuang, I. L., L. M. K. Vandersypen, X. Zhou, D. W. Leung, and S. Lloyd. 1998. Experimental Realization of a Quantum Algorithm. *Nature* **393**: 143.
- Cory, D. G., A. F. Fahmy, and T. F. Havel. 1997. Ensemble Quantum Computing by NMR-Spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **94**: 1634.
- Cory, D. G., R. Laflamme, E. Knill, L. Viola, T. F. Havel, N. Boulant, et al. 2000. NMR Based Quantum Information Processing: Achievements and Prospects. *Fortschr. Phys.* **48**: 875.
- Cory, D. G., W. Maas, M. Price, E. Knill, R. Laflamme, W. H. Zurek, et al. 1998. Experimental Quantum Error Correction. *Phys. Rev. Lett.* **81**: 2152.
- Cummins, H. K., and J. A. Jones. 1999. Use of Composite Rotations to Correct Systematic Errors in NMR Quantum Computation. [Online]: <http://eprints.lanl.gov> (quant-ph/9911072).
- DiVincenzo, D. P. 1995. Two-Bit Gates are Universal for Quantum Computation. *Phys. Rev. A* **51**: 1015.

Raymond Laflamme graduated from the University of Laval, Canada, in 1983 with a Bachelor's degree in physics. Raymond received a Ph.D. in applied mathematics and theoretical physics from the University of Cambridge, England, where he worked under the direction of Stephen Hawking.



Raymond then became a Killam post-doctoral fellow at the University of British Columbia before returning to Cambridge as a research fellow at Peterhouse College. In 1992, he came to Los Alamos as a Director-funded postdoctoral fellow, became an Oppenheimer fellow in 1994, and a technical staff member in 1997. In the fall of 2001, he moved to the newly created Perimeter Institute at the University of Waterloo, where he currently holds a Canadian Research Chair in Quantum Information. His research includes both theoretical and experimental investigations of quantum information processing.

David Cory received both his undergraduate and graduate degrees in chemistry from Case Western Reserve University. He is currently the Rasmussen Professor of Nuclear Engineering at the Massachusetts Institute of Technology. David and his students work to advance the instrumentation, methodology, and applications of nuclear magnetic resonance spectroscopy.



Contact Information**R. Laflamme:** laflamme@iqc.ca**E. Knill:** knill@lanl.gov**D. Cory:** dcory@mit.edu**E. M. Fortunato:** evanmf@mit.edu**T. Havel:** efhave1@mit.edu**C. Miquel:** miquel@df.uba.ar**R. Martinez:** rudy@lanl.gov**C. Negrevergne:** cjn@lanl.gov**G. Ortiz:** g_ortiz@lanl.gov**M. A. Pravia:** praviam@mit.edu**Y. Sharf:** ysharf@mit.edu**S. Sinha:** suddha@mit.edu**R. Somma:** somma@lanl.gov**L. Viola:** lviola@lanl.gov

- Ernst, R. R., G. Bodenhausen, and A. Wokaun. 1994. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Oxford: Oxford University Press.
- Gershenfeld, N. A., and I. L. Chuang. 1997. Bulk Spin Resonance Quantum Computation. *Science* **275**: 350.
- Glaser, S. J., T. Schulte-Herbruggen, M. Sieveking, O. Schedletsky, N. C. Nielsen, O. W. Sørensen, and C. Griesinger. 1998. Unitary Control in Quantum Ensembles: Maximizing Signal Intensity in Coherent Spectroscopy. *Science* **280**: 421.
- Hahn, E. L. 1950. Spin Echoes. *Phys. Rev.* **80**: 580.
- Jones, J. A., and E. Knill. 1999. Efficient Refocussing of One Spin and Two Spin Interactions for NMR. *J. Magn. Reson.* **141**: 322.
- Jones, J. A., M. Mosca, and R. H. Hansen. 1998. Implementation of a Quantum Search Algorithm on a Quantum Computer. *Nature* **392**: 344.
- Kane, B. E. 1998. A Silicon-Based Nuclear Spin Quantum Computer. *Nature* **393**: 133.
- Knill, E., and R. Laflamme. 1998. On the Power of One Bit of Quantum Information. *Phys. Rev. Lett.* **81**: 5672.
- Knill, E., I. L. Chuang, and R. Laflamme. 1998. Effective Pure States for Bulk Quantum Computation. *Phys. Rev. A* **57**: 3348.
- Knill, E., R. Laflamme, R. Martinez, and C. Negrevergne. 2001. Implementation of the Five Qubit Error Correction Benchmark. *Phys. Rev. Lett.* **86**: 5811.
- Knill, E., R. Laflamme, R. Martinez, and C.-H. Tseng. 2000. An Algorithmic Benchmark for Quantum Information Processing. *Nature* **404**: 368.
- Leung, D. W., I. L. Chuang, F. Yamaguchi, and Y. Yamamoto. 1999. Efficient Implementation of Selective Recoupling in Heteronuclear Spin Systems Using Hadamard Matrices. [Online]: <http://eprints.lanl.gov/quant-ph/9904100>.
- Leung, D., L. Vandersypen, X. L. Zhou, M. Sherwood, C. Yannoni, M. Kubinec, and I. L. Chuang. 1999. Experimental Realization of a Two-Bit Phase Damping Quantum Code. *Phys. Rev. A* **60**: 1924.
- Levitt, M. H. 1982. Symmetrical Composite Pulse Sequences for NMR Population-Inversion 1. Compensation for Radiofrequency Field Inhomogeneity. *J. Magn. Reson.* **48**: 234.
- Lloyd, S. 1995. Almost Any Quantum Logic Gate is Universal. *Phys. Rev. Lett.* **75**: 346.
- Lloyd, S. 1996. Universal Quantum Simulators. *Science* **273**: 1073.
- Mansfield, P., and P. Morris. 1982. NMR Imaging in Biomedicine. *Adv. Magn. Reson.* **S2**: 1.
- Miquel, C., J. P. Paz, M. Saraceno, B. Knill, R. Laflamme, and C. Negrevergne. 2002. Interpretation of Tomography and Spectroscopy as Dual Forms of Quantum Computations. *Nature* **418**: 59.
- Purcell, E. M., H. C. Torrey, and R. V. Pound. 1946. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev.* **69**: 37.
- Sackett, C. A., D. Kielpinski, B. B. King, C. Langer, V. Meyer, C. J. Myatt, et al. 2000. Experimental Entanglement of Four Particles. *Nature* **404**: 256.
- Schulman, L. J., and U. Vazirani. 1998. Scalable NMR Quantum Computation. In *Proceedings of the 31th Annual ACM Symposium on the Theory of Computation (STOC)*. p. 322. El Paso, TX: ACM Press.
- Sharf, Y., T. F. Havel, and D. G. Cory. 2000. Spatially Encoded Pseudopure States for NMR Quantum-Information Processing. *Phys. Rev. A* **62**: 052314.
- Sharf, Y., D. G. Cory, S. S. Somaroo, E. Knill, R. Laflamme, W. H. Zurek, and T. F. Havel. 2000. A Study of Quantum Error Correction by Geometric Algebra and Liquid-State NMR Spectroscopy. *Mol. Phys.* **98**: 1347.
- Somma, R., G. Ortiz, J. E. Gubernatis, R. Laflamme, and E. Knill. 2002. Simulating Physical Phenomena by Quantum Networks. *Phys. Rev. A* **65**: 042323
- Sørensen, O. W., G. W. Eich, M. H. Levitt, G. Bodenhausen, and R. R. Ernst. 1983. Product Operator-Formalism for the Description of NMR Pulse Experiments. *Prog. Nucl. Magn. Reson. Spectrosc.* **16**: 163.
- Stoll, M. E., A. J. Vega, and R. W. Vaughan. 1977. Explicit Demonstration of Spinor Character for a Spin-1/2 Nucleus Using NMR Interferometry. *Phys. Rev. A* **16**: 1521.
- Viola, L., and S. Lloyd. 1998. Dynamical Suppression of Decoherence in Two-State Quantum Systems. *Phys. Rev. A* **58**: 2733.
- Viola, L., E. Knill, and S. Lloyd. 1999. Dynamical Decoupling of Open Quantum Systems. *Phys. Rev. Lett.* **82**: 2417.
- Viola, L., E. M. Fortunato, M. A. Pravia, E. Knill, R. Laflamme, and D. G. Cory. 2001. Experimental Realization of Noiseless Subsystems for Quantum Information Processing. *Science* **293**: 2059.
- Warren, W. S., H. Rabitz, and M. Dahleh. 1993. Coherent Control of Quantum Dynamics: The Dream is Alive. *Science* **259**: 1581.
- Zalta, E. N., ed. 2002. *Stanford Encyclopedia of Philosophy*. Stanford University, Stanford, CA: The Metaphysics Research Lab.

Glossary

- Bloch sphere.** A representation of the state space of a qubit using the unit sphere in three dimensions. See Figure 3 in the main text of the article.
- Crosstalk.** In using physical control to implement a gate, crosstalk refers to unintended effects on qubits not involved in the gate.
- Decoupling.** A method for turning off the interactions between two sets of spins. In NMR, this task can be achieved if one applies a rapid sequence of refocusing pulses to one set of spins. The other set of spins can then be controlled and observed as if independent of the first set.
- Deviation of a state.** If ρ is a density matrix for a state and $\rho = \alpha\mathbb{1} + \beta\sigma$, then σ is a deviation of ρ .
- Ensemble computation.** Computation with a large ensemble of identical and independent computers. Each step of the computation is applied identically to the computers. At the end of the computation, the answer is determined from a noisy measurement of the fraction p_1 of the computers whose answer is “1.” The amount of noise is important for resource accounting: To reduce the noise to below ε requires increasing the resources used by a factor of the order of $1/\varepsilon^2$.
- Equilibrium state.** The state of a quantum system in equilibrium with its environment. In the present context, the environment behaves like a heat bath at temperature T , and the equilibrium state can be written as $\rho = e^{-H/kT}/Z$, where H is the effective internal Hamiltonian of the system and Z is determined by the identity $\text{tr}\rho = 1$.
- FID.** Free induction decay. To obtain a spectrum on an NMR spectrometer after having applied pulses to a sample, one measures the decaying planar magnetization induced by the nuclear spins as they precess. The x - and y -components $M_x(t)$ and $M_y(t)$ of the magnetization as a function of time are combined to form a complex signal $M(t) = M_x(t) + iM_y(t)$. The record of $M(t)$ over time is called the FID, which is Fourier-transformed to yield the spectrum.
- Inversion.** A pulse that flips the z -component of the spin. Note that any 180° rotation around an axis in the xy -plane has this effect.
- J -coupling.** The type of coupling present between two nuclear spins in a molecule in the liquid state.
- Labeled molecule.** A molecule in which some of the nuclei are substituted by less common isotopes. A common labeling for NMR QIP involves replacing the naturally abundant carbon isotope ^{12}C , with the spin-1/2 isotope ^{13}C .
- Larmor frequency.** The precession frequency of a nuclear spin in a magnetic field. It depends linearly on the spin’s magnetic moment and the strength of the field.
- Logical frame.** The current frame with respect to which the state of a qubit carried by a spin is defined. There is an absolute (laboratory) frame associated with the spin observables σ_x , σ_y , and σ_z . The observables are spatially meaningful. For example, the magnetization induced along the x -axis is proportional to $\text{tr}(\sigma_x|\psi\rangle\langle\psi|)$, where $|\psi\rangle$ is the physical state of the spin. Suppose that the logical frame is obtained from the physical frame by a rotation by an angle of θ around the z -axis. The observables for the qubit are then given by $\sigma_x^{(L)} = \cos(\theta)\sigma_x + \sin(\theta)\sigma_y$, $\sigma_y^{(L)} = \cos(\theta)\sigma_y - \sin(\theta)\sigma_x$, and $\sigma_z^{(L)} = \sigma_z$. As a result, the change to the logical frame transforms the physical state to a logical state according to $|\phi\rangle_L = e^{i\sigma_z\theta/2}|\psi\rangle$. That is, the logical state is obtained from the physical state by a $-\theta$ rotation around the z -axis. A resonant logical frame is used in NMR to compensate for the precession induced by the strong external field.
- Magnetization.** The magnetic field induced by an ensemble of magnetic spins. The magnitude of the magnetization depends on the number of spins, the extent of alignment, and the magnetic moments.

Nuclear magnetic moment. The magnetic moment of a nucleus determines the strength of the interaction between its nuclear spin and a magnetic field. The precession frequency ω of a spin-1/2 nucleus is given by μB , where μ is the nuclear magnetic moment and B the magnetic field strength. For example, for a proton, $\mu = 42.7$ Mhz/T.

NMR spectrometer. The equipment used to apply rf pulses to and observe precessing magnetization from nuclear spins. Typical spectrometers consist of a strong, cylindrical magnet with a central bore in which there is a “probe” that contains coils and a sample holder. The probe is connected to electronic equipment for applying rf currents to the coils and for detecting weak oscillating currents induced by the nuclear magnetization.

Nuclear spin. The quantum spin degree of freedom of a nucleus. It is characterized by its total spin quantum number, which is a multiple of 1/2. Nuclear spins with spin 1/2 are two-state quantum systems and can therefore be used as qubits immediately.

Nutation. The motion of a spin in a strong z -axis field caused by a resonant pulse.

Nutation frequency. The angular rate at which a resonant pulse causes nutation of a precessing spin around an axis in the plane.

One-qubit quantum computing. The model of computation in which one can initialize any number of qubits in the state where Qubit 1 is in the state $|0\rangle_1$ and all the other qubits are in a random state. One can then apply one- and two-qubit unitary quantum gates and make one final measurement of the state of Qubit 1 after which the system is reinitialized. The model can be used to determine properties of the spectral density function of a Hamiltonian, which can be emulated by a quantum computer (Knill and Laflamme 1998).

Peak group. The spectrum of an isolated nuclear spin consists of one peak at its precession frequency. If the nuclear spin is coupled to others, this peak “splits,” and multiple peaks are observed near the precession frequency. The nuclear spin’s peak group consists of these peaks.

Precession. An isolated nuclear spin’s state can be associated with a spatial direction with the help of the Bloch sphere representation. If the direction rotates around the z -axis at a constant rate, we say that it precesses around the z -axis. The motion corresponds to that of a classical top experiencing a torque perpendicular to both the z -axis and the spin axis. For a nuclear spin, the torque can be caused by a magnetic field along the z -axis.

Projective measurement. A measurement of a quantum system determined by a complete set of orthogonal projections whose effect is to apply one of the projections to the system (“wave function collapse”) with a probability determined by the amplitude squared of the projected state. Which projection occurred is known after the measurement. The simplest example is that of measuring Qubit q in the logical basis. In this case, there are two projections, namely, $P_0 = |0\rangle_q \langle 0|$ and $P_1 = |1\rangle_q \langle 1|$. If the initial state of all the qubits is $|\psi\rangle$, then the probabilities of the two measurement outcomes 0 and 1 are $p_0 = \langle \psi | P_0 | \psi \rangle$ and $p_1 = \langle \psi | P_1 | \psi \rangle$, respectively. The state after the measurement is $P_0 |\psi\rangle / \sqrt{p_0}$ for outcome 0 and $P_1 |\psi\rangle / \sqrt{p_1}$ for outcome 1.

Pseudopure state. A state with deviation given by a pure state $|\psi\rangle \langle \psi|$.

Pulse. A transient field applied to a quantum system. In the case of NMR QIP, pulses are rotating magnetic fields (rf pulses) whose effects are designed to cause specific rotations of the qubit states carried by the nuclear spins.

Radio-frequency (rf) pulse. A pulse resonant at radio frequencies. Typical frequencies used in NMR are in this range.

Refocusing pulse. A pulse that causes a 180° rotation around an axis in the plane.

A typical example of such a rotation is $e^{-i\sigma_x\pi/2} = -i\sigma_x$, which is a 180° x -rotation.

Resonant rf pulse. A pulse whose field oscillates at the same frequency as the precession frequency of a target nuclear spin. Ideally, the field is in the plane, rotating at the same frequency and in the same direction as the precession. However, as long as the pulse field is weak compared with the precession frequency (that is, by comparison, its nutation frequency is small), the nuclear spin is affected only by the corotating component of the field. As a result, other planar components can be neglected, and a field oscillating in a constant direction in the plane has the same effect as an ideal resonant field.

Rotating frame. A frame rotating at the same frequency as the precession frequency of a spin.

Rotation. In the context of spins and qubits, a rotation around σ_u by an angle θ is an operation of the form $e^{-i\sigma_u\theta/2}$. The operator σ_u may be any unit combination of Pauli matrices that defines an axis in three space. In the Bloch sphere representation, the operation has the effect suggested by the word “rotation.”

Spectrum. In the context of NMR, the Fourier transform of an FID.

Weak measurement. A measurement involving only a weak interaction with the measured quantum system. Typically, the measurement is ineffective unless an ensemble of these quantum systems is available so that the effects of the interaction add up to a signal detectable above the noise. The measurement of nuclear magnetization used in NMR is weak in this sense.

Realizing a Noiseless Subsystem in an NMR Quantum Information Processor

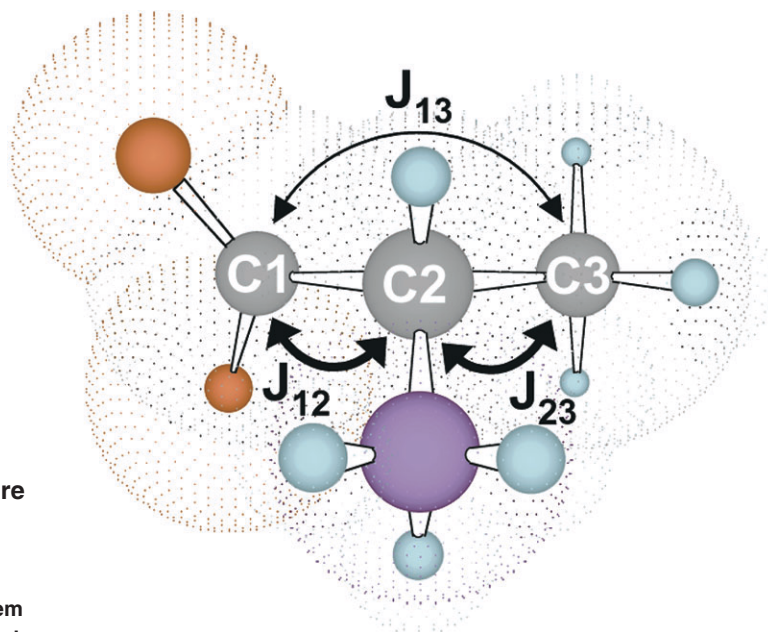


Figure 1. Molecular Structure of ¹³C-Labeled Alanine

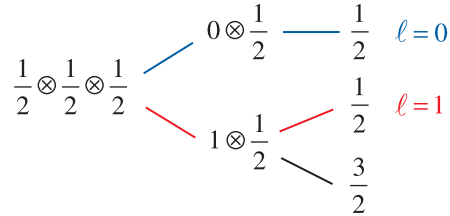
The diagram shows the three carbon-13 (¹³C) spins used as qubits in the noiseless subsystem experiment as well as the relevant J -couplings between those qubits.

Lorenza Viola and Evan M. Fortunato

The fact that the occurrence of symmetries in a physical system generally implies the existence of conserved quantities and that these symmetries can be exploited to ease the understanding of the system's behavior is a well-known lesson in physics. The notion of a noiseless subsystem (NS) (Knill et al. 2000) captures this lesson in the context of quantum information processing (QIP), where the challenge is to protect information against the detrimental effects of noise. The link between symmetries, conserved quantities, and NS was discussed at length on page 216 of the article "Introduction to Error Correction". The essential message is that, by encoding information into an abstract subsystem that corresponds to a preserved degree of freedom, noiselessness is guaranteed even if errors still evolve the overall system's state.

Here, we focus on the NS of three spin-1/2 particles introduced in the above-mentioned article (see page 201), along with a discussion of the error-correcting properties of this NS. The physical system is composed of three qubits, subjected to a "far-field" interaction with the environment, whereby the latter couples to the qubits without distinguishing among them. The resulting collective-noise model involves all possible error operators that are symmetric under permutation of the three particles and is specified in terms of the error generators $J_u = (\sigma_u^{(1)} + \sigma_u^{(2)} + \sigma_u^{(3)})/2$, where $u = x, y, z$. By recalling the meaning of the single-spin Pauli operators σ_u^k , the observable J_u represents the projection of the total spin angular momentum \mathbf{J} along the u -axis. Because the total-spin observable $J^2 = \mathbf{J} \cdot \mathbf{J}$ commutes with the error generators and z defines the quantization axis, the eigenvalues j and j_z of J^2 and J_z , respectively, provide useful quantum numbers to label basis states for the three particles.

The NS of interest resides in the four-dimensional subspace $\mathcal{H}_{1/2}$ of the states carrying total angular momentum $j = 1/2$ and having a total z -component $j_z = \pm 1/2$. However, specifying j and j_z does not suffice for completely labeling the states in $\mathcal{H}_{1/2}$: An additional quantum number is needed for removing the two-dimensional degeneracy that remains. Physically, this degeneracy simply means that there are two distinct paths for obtaining a total angular momentum $j = 1/2$ out of three elementary $1/2$ angular momenta:



Let the additional quantum number $\ell = 0, 1$ label the two possible routes in the above diagram. Because collective noise does not distinguish among the individual spins and the final eigenvalue j is the same for both paths, the noise can neither distinguish the realized value of ℓ nor change that value. This conserved quantum number can be directly related to the eigenvalues $s_z = \pm 1$ of the $\sigma_z^{(\text{NS})}$ observable of a noiseless qubit, $s_z = 2\ell - 1$. In general, noiseless qubit operators will remain invariant under rotations. The simplest scalars under the rotations are the dot products $s_{12} = \sigma^{(1)} \cdot \sigma^{(2)}$, $s_{23} = \sigma^{(2)} \cdot \sigma^{(3)}$, and $s_{31} = \sigma^{(3)} \cdot \sigma^{(1)}$.

Thus, $\sigma_u^{(\text{NS})}$ observables for the noiseless qubit, where $u = x, y, z$, can be constructed by combining s_{12} , s_{23} , s_{31} , and the identity into three operators that “behave like” the Pauli matrices (Viola et al. 2001a). A good choice is given by $\sigma_x^{(\text{NS})} = 1/2(1 + s_{23})$, $\sigma_y^{(\text{NS})} = \sqrt{3}/6(s_{31} - s_{12})$, and $\sigma_z^{(\text{NS})} = i\sigma_y^{(\text{NS})}\sigma_x^{(\text{NS})}$, where projection onto the relevant $\mathcal{H}_{1/2}$ subspace is understood. Note that the action corresponding to $\sigma_x^{(\text{NS})}$ is simply a permutation exchanging the last two spins. (For an alternative construction of the NS observables, see the article “Introduction to Error Correction,” page 216.) Identifying the NS through its observables is equivalent to identifying it through the explicit state space correspondence given in Equation (28) of the above-mentioned article.

The experimental implementation of the three-qubit NS (Viola et al. 2001b) was performed with liquid-state NMR techniques. The three spin-1/2 carbon nuclei of carbon-13-labeled alanine were used as qubits (Figure 1). The information to protect is an arbitrary one-qubit state, $|\psi\rangle = a|0\rangle + b|1\rangle$, where a and b are arbitrary complex amplitudes, and $\langle\psi|\psi\rangle = 1$. This information is initially stored in spin 3, meaning that the three carbon spins are initialized in a pseudopure state corresponding to $|0\rangle_1|0\rangle_2|\psi\rangle_3 = |00\rangle\psi = a|000\rangle + b|001\rangle$. A unitary transformation U_{enc} encodes this input state into a superposition of the two basis states in $\mathcal{H}_{1/2}$ with $j = 1/2$ and $j_z = -1/2$. That is,

$$U_{\text{enc}}|00\rangle\psi \leftrightarrow a|\downarrow\rangle \cdot |0\rangle + b|\downarrow\rangle \cdot |1\rangle = |\downarrow\rangle \cdot |\psi\rangle, \quad (1)$$

where the subsystem representation of Equation (28) has been used.

The three qubits remain stored in the NS memory for a fixed evolution period t_{ev} , during which errors can occur. In a given set of experiments, these errors are designed to implement a desired collective-noise process $\mathcal{E}_{\text{coll}}$ described by a set of error operators $\{E_a\}$. Because of their collective nature, these errors affect only the syndrome subsystem in the pair. Finally, following the evolution period, the unitary transformation U_{dec} decodes a generic noisy state $E_a(|\downarrow\rangle \cdot |\psi\rangle)$ in $\mathcal{H}_{1/2}$ back to the computational basis.

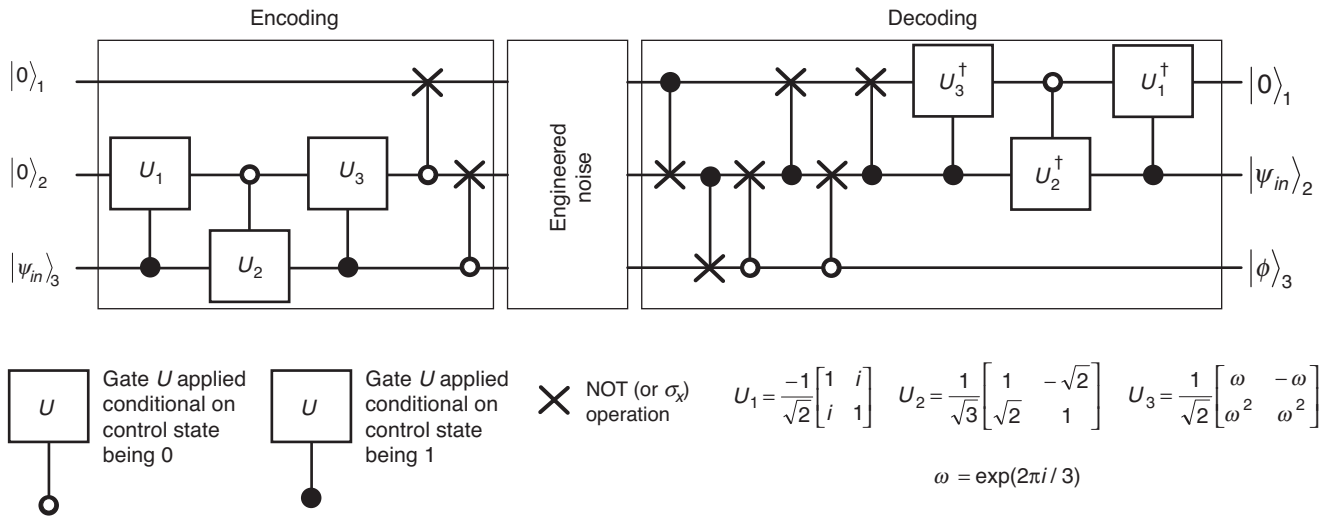


Figure 2. Logical Quantum Network

The diagram shows a logical quantum network for the three-qubit NS experiment. The logical manipulations were translated into sequences of radio-frequency pulses and delays, and complete pulse programs for U_{enc} and U_{dec} resulted from the compilation of the partial pulse programs for individual gates. The pulses were designed to ensure self-refocusing of all the unwanted J -coupling and chemical-shift evolutions.

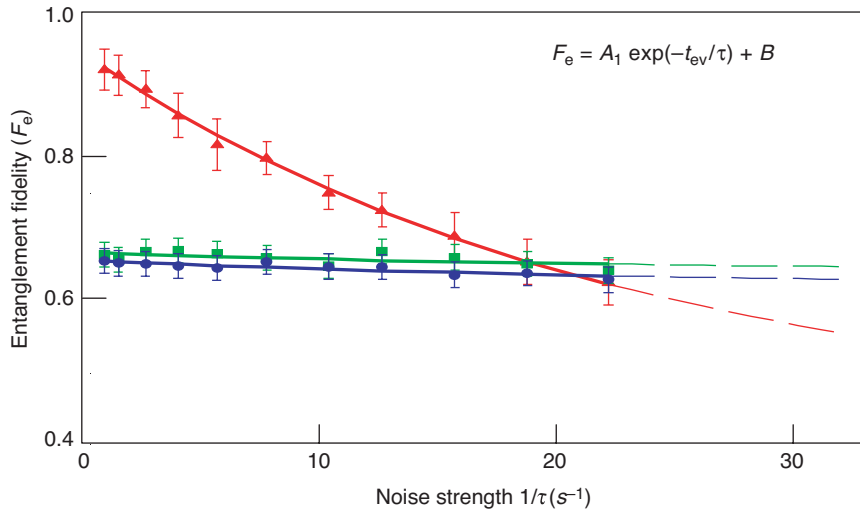
This procedure has the effect of returning the quantum state $|\psi\rangle$ onto qubit 2 upon discarding (“tracing over”) spins 1 and 3,

$$\text{Tr}_{1,3}\{U_{dec}[\mathcal{E}_{coll}(|\downarrow\rangle\langle\downarrow| \cdot |\psi\rangle\langle\psi|)]U_{enc}^{-1}\} = |\psi\rangle\langle\psi| \quad (2)$$

Figure 2 is a sketch of the quantum network for the experiment.

During the delay period between encoding and decoding, we use gradient diffusion techniques to engineer a desired collective-noise process. In order to fully explore the robustness properties of information encoded in the NS, we applied various error models corresponding to noise along a single axis (see Figure 3), as well as more complicated double- and triple-axis noise processes obtained by “cascading” the action of error models along different spatial directions, in sequence, within a single evolution period (see Table I). To quantify the accuracy of the implemented NS in preserving the quantum data $|\psi\rangle$, we experimentally extracted the entanglement fidelity F_e of the overall process (including encoding, decoding, and engineered noise during storage), where $F_e = 1$ implies perfect preservation.

Our results in Figure 3 and Table I indicate that, as expected, the effects of the applied noise increase exponentially as a function of noise strength for unencoded (UN) information but are largely independent of noise strength for information encoded in the NS. That independence demonstrates that the NS functions as an “infinite-distance” quantum error-correcting code for arbitrary collective errors. On the other hand, the F_e is always about the same and less than 1 in all the NS experiments. The constant reduction in fidelity is suggestive of errors introduced during encoding and decoding manipulations, as well as of noise due to natural noncollective relaxation processes during the whole experiment. ■



$F_e = A_1 \exp(-t_{ev}/t) + B$		Quantum Process
A_1	B	
0.51 ± 0.04	0.43 ± 0.03	▲ UN, y-axis noise
0.03 ± 0.03	0.64 ± 0.02	■ NS-encoded, y-axis noise
0.03 ± 0.03	0.62 ± 0.02	● NS-encoded, z-axis noise

Table I. Entanglement Fidelities for Engineered Collective Noise along Two and Three Axes

Quantum Process	Entanglement Fidelity (F_e)
Q_{zx}^{UN}	0.24
Q_{00}^{NS}	0.70
Q_{zx}^{NS}	0.70
Q_{zy}^{NS}	0.70
Q_{000}^{NS}	0.67
Q_{yzx}^{NS}	0.66

Q stands for the one-qubit processes implemented during each run.

Superscripts tell whether the system has been encoded or not.

Subscripts zx, zy, and yzx are for the axes along which noise processes with maximum achievable strength were applied in cascade. Subscripts 00 and 000 indicate that no noise was applied.

Two subscripts indicate shorter delay periods than three subscripts.

Statistical uncertainties in all F_e values are approximately 2%.

Further Reading

- Knill, E., R. Laflamme, and L. Viola. 2000. Theory of Quantum Error Correction for General Noise. *Phys. Rev. Lett.* **84**: 2525.
- Viola, L., E. Knill, and R. Laflamme. 2001a. Constructing Qubits in Physical Systems. *J. Phys. A* **34** (35): 7067.
- Viola L., E. M. Fortunato, M. A. Pravia, E. Knill, R. Laflamme, and D. G. Cory. 2001b. Experimental Realization of Noiseless Subsystems for Quantum Information Processing. *Science* **293**: 2059.

Figure 3. Entanglement Fidelities for Engineered Collective Noise along a Single Axis

The fidelity of UN information subjected to engineered collective noise along the y-axis (red) decreases exponentially with noise strength τ^{-1} whereas the fidelity of NS-encoded information subjected to collective noise along either the y-axis (green) or the z-axis (black) remains almost constant independent of noise strength.

In each case, noise was applied for a fixed evolution period t_{ev} of approximately 44 ms.

The flatness of the curve interpolating the NS data demonstrates the behavior of the NS as an infinite-distance quantum error-correcting code for single-axis collective errors of arbitrary strength. The smooth fits to the data are derived from the exponential and parameters displayed under the figure.

Lorenza Viola obtained a Ph.D. degree in physics from the University of Padova (Italy) in 1996. After being a postdoctoral fellow at the D'Arbeloff Laboratory



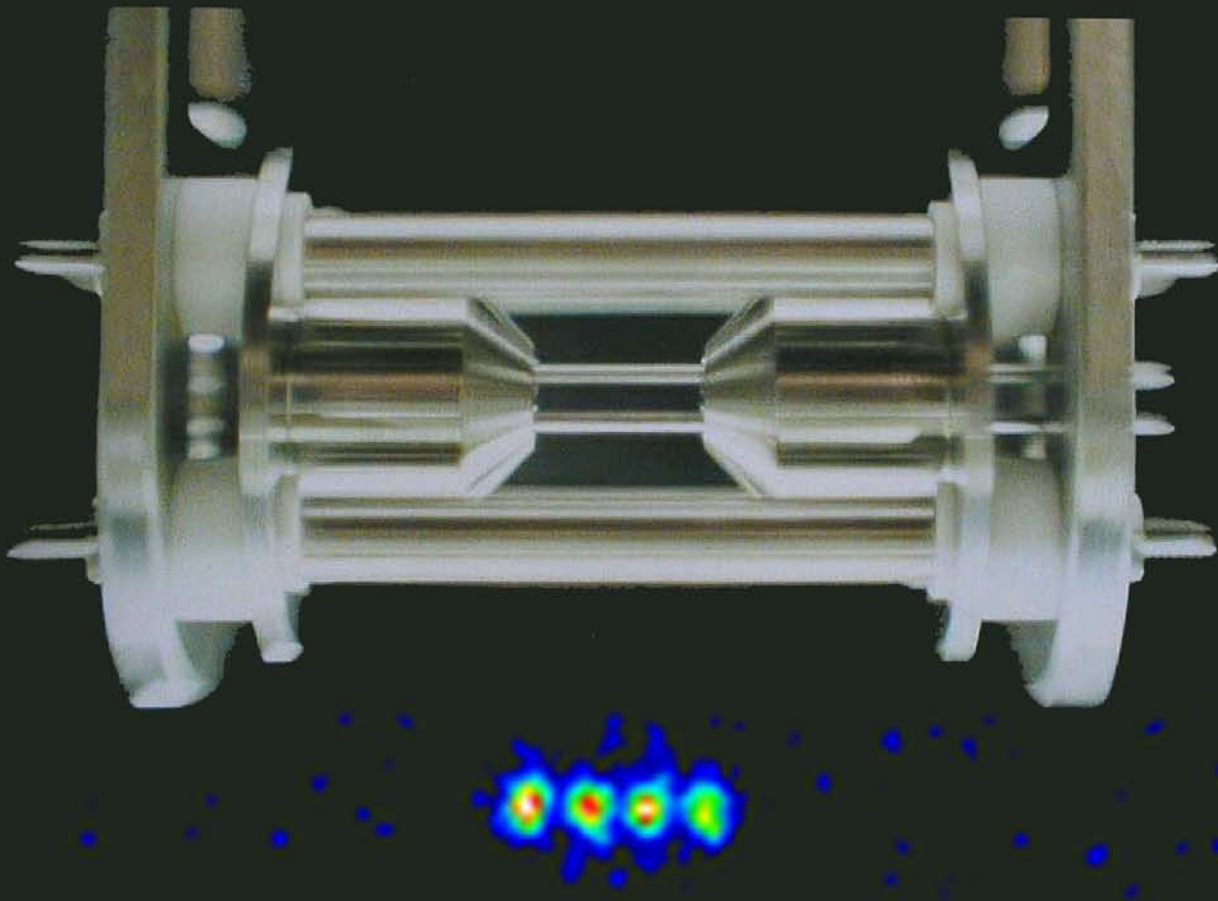
at the Massachusetts Institute of Technology, Lorenza became a Director-funded and then a J. R. Oppenheimer postdoctoral fellow at Los Alamos. Her recent research has been focused on quantum information science, with emphasis on devising schemes for controlling noisy quantum devices and performing robust quantum computation. Lorenza has been a key contributor to the development of quantum error-suppression techniques based on dynamical decoupling and to the theoretical characterization and experimental verification of the notion of a noiseless subsystem as the most general approach to noise-free information storage.

“...it seems that the laws of physics present no barrier to reducing the size of computers until bits are the size of atoms, and quantum behavior holds dominant sway.”

—R. P. Feynman, 1985

Ion-Trap Quantum Computation

Michael H. Holzscheiter



Quantum computation requires a very special physical environment. Numerous operations must be performed on the quantum states of the qubits (or quantum bits) before those states decohere, or lose the interlocking phase relationships that give quantum computation

its power. Thought to be an unavoidable outcome of the interaction between the quantum state and the environment, decoherence threatens the life of a quantum system.

Any attempt at building a real quantum computer therefore leads to what some scientists refer to as the

yin-yang of quantum computation: On the one hand, the qubits must interact weakly with the environment in order to limit decoherence. On the other hand, they must be easily accessible from the outside and must interact strongly with each other, or else we could not manipulate the

quantum state, implement quantum algorithms, and read out the result of a calculation in a timely fashion. How can we hope to meet such contradictory goals?

Ion-trap quantum computers, as originally proposed by Ignacio Cirac and Peter Zoller (1995), offer a possible solution to this dilemma. As its name implies, an ion trap confines charged particles to a definite region of space with magnetic and electric fields. In a specific realization of such a trap, called a linear radio-frequency quadrupole (RFQ) trap, or a linear Paul trap (Raizen et al. 1992, Walther 1994), time-varying electric fields are used to hold a line of ions in place—like pearls on a string. These ions serve as the physical qubits of the quantum computer. Immobilized by the trapping fields and confined inside an ultrahigh vacuum chamber, they are effectively isolated from the environment. However, by addressing individual ions with sharply defined laser beams, we can initialize the computer, control the qubit states during the operation of logic gates, and read out the results at the end of the computation. The interaction between the individual ions is mediated by the Coulomb force between the charged particles.

This article discusses the design principles for isolating single ions in a linear Paul trap (Paul et al. 1958), whose operational principles are described in detail. The individual elements of an ion-trap computer will be introduced, and how to initialize, manipulate, and interrogate the qubits will be explained. Specific schemes that were implemented in the quantum computation project at Los Alamos (Hughes et al. 1998) will illustrate the descriptions. Ion-trap quantum computation is rapidly evolving, and numerous groups around the world are developing new ideas and experimental techniques. The reader will get a flavor of this activity in the last section of the article, which summarizes sev-

eral important results achieved within the last few years: the on-demand creation of entangled states of up to four ions by the National Institute of Science and Technology (NIST) group in Boulder, Colorado; the development of a novel cooling scheme by a group at the University of Innsbruck, Austria, which would allow researchers to quickly cool large numbers of trapped ions with drastically reduced operational overhead; and the construction of an effective defense against the forces of decoherence.

The Physics of Ion Traps

Two basic types of devices can confine charged particles to well-defined regions of free space: Penning traps and Paul traps. The Penning trap, which was primarily developed by Hans Dehmelt at the University of Washington in Seattle, uses a strong magnetic field and a static electric field to create a nearly perfect three-dimensional, harmonic trapping potential (Dehmelt 1967). Some of the most precise tests of fundamental physical symmetries to date have been conducted with this device, whose operating principles are described in the box “The Penning Trap” on the next two pages.

Although Penning traps nicely solve the fundamental problems of ion confinement, so far they have not been used for quantum computation. The trap’s strong magnetic field causes ions to move rapidly in a circle (the cyclotron motion discussed in the box), whereas we want the physical qubits to have as little motion as possible. That is why the favored trap for quantum computation is the Paul trap, in which there is no magnetic field and oscillating electric fields (as opposed to static ones) confine the ions. This device was invented by Wolfgang Paul from the University of Bonn in Germany,

who shared the 1989 Nobel Prize in physics with Dehmelt.

Paul enjoyed telling the following anecdote about how he hit upon the idea for his device. Germans like soft-boiled eggs for breakfast, and on a particular Sunday morning, Paul had prepared two eggs of different sizes and had placed them on a serving tray. When he started to walk, tray in hand, toward the bedroom to surprise his wife with breakfast in bed, the eggs began to roll. He counteracted their motion by shaking the tray and was able to confine the larger egg to the center by shaking with a particular frequency and amplitude. (It was certainly not a well-defined harmonic shaking.) The smaller egg, however, kept rolling toward the edge, so Paul changed amplitude and frequency and successfully prevented this egg from falling, at the expense of allowing the larger one to wobble toward the edge. Whether he ever reached the bedroom with both eggs on the tray and enjoyed a leisurely breakfast with his wife remains unknown, but that morning Paul realized not only the basic principle of the RFQ trap but also the mass-selective feature of such an instrument. At that time, he was keen on developing a mass filter for ions, that is, a two-dimensional structure that could transmit an ion with a specific charge-to-mass ratio and not any other ratio. Eventually, Paul’s idea was used to generate three-dimensional, mass-selective confinement systems, but Cirac and Zoller returned to the original two-dimensional structure and proposed using it as the basis for a quantum computer.

The Linear Paul Trap. To understand the linear RFQ trap, consider a positively charged ion floating in free space and surrounded by four infinitely long conducting rods, as shown in Figure 1. We can give one pair of opposing rods a positive charge and the other pair a negative charge

Continued on page 268

The Penning Trap

Decades before individual ions were considered as candidates for qubits in a quantum computer, experimental physicists were challenged to realize a simpler Gedanken, or thought, experiment embodied by the statement often made by theorists: “Consider a single (silver) ion in a uniform magnetic field” (Tanoudji et al. 1977). Thought became reality in 1973, when Hans Dehmelt and his colleagues at the University of Washington in Seattle were able to capture a single charged particle in a Penning trap. The ion that drifted into the central region of that device was trapped by a strong, uniform magnetic field and by the electrostatic field produced by a set of specially shaped electrodes. The entire device operated under ultrahigh vacuum to limit the interactions between the ion and the background atoms.

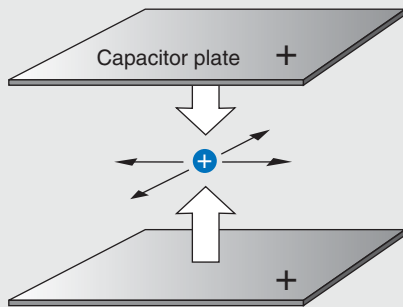


Figure A. Electrostatic Forces
The positively charged particle is repulsed by the capacitor plates but is free to move anywhere in the horizontal plane.

The University of Washington group refined the technique and used the trap to confine a single electron (Wineland et al. 1973) and later a single barium ion, using an RFQ Paul trap (Neuhauser et al. 1980), and performed precision spectroscopy on these systems. The special arrangement of fields caused the single electron to behave as if it were bound to a nucleus for it displayed a set of energy levels, or excited states, similar to those of the hydrogen atom. Dehmelt therefore named his electron in a Penning trap “geonium—a single electron bound to Earth.” The artificial geonium “atom” was, in a sense, closer to perfection than a real atom. The spacing between energy levels was nearly constant because it reflected the trap’s nearly perfect harmonic-oscillator potential. Dehmelt and coworkers used geonium to perform some of the most precise tests of fundamental symmetries. In a more mundane fashion, Dehmelt called the ion ASTRID (for “a single trapped ion dancing”). (Perhaps, if you keep an ion or electron for such a long time, you may become attached to it.)

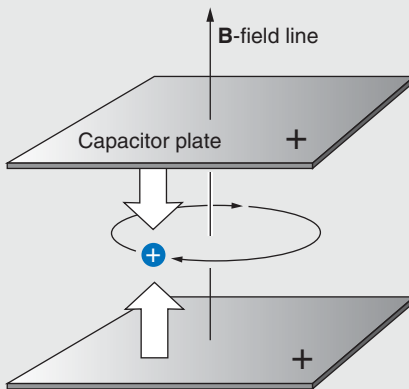


Figure B. Applying a Magnetic Field
A magnetic field causes the ion to circle around a field line (cyclotron motion), thus confining the ion in the horizontal plane.

To understand the operating principles of the Penning trap, consider a charged particle (ion) moving freely in space. To confine it to a specific spot in space, we can apply electrical forces to its charge. If we place the ion between two parallel conducting plates that are charged to an electric potential of the same sign as the ion, the Coulomb repulsion will keep the particle from moving closer to either plate (see Figure A).

The ion can still move in directions parallel to the conductors. We can try to remove all escape routes by placing more conductors around the particle. But Michael Faraday discovered more than 150 years ago that an electric field cannot penetrate a closed metal enclosure—hence, the penchant for science museums to place a person inside a “Faraday cage” that is then exposed to violent lightning bolts. The courageous volunteer remains unharmed because the lightning’s electric field vanishes inside the cage. Similarly, if we fully enclose our particle in a cage of conducting plates, the electric field disappears, and we lose the forces holding the particle from the walls.

A more successful approach is to use the fact that an electric charge moving in a magnetic field will experience a force perpendicular to the direction of both the magnetic field and the particle’s velocity (the Lorentz force $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$). Therefore, if we apply a magnetic field perpendicular to our parallel conducting plates (see Figure B), we force the ion onto a circular path around the magnetic field line, closing off the sideway escape routes.

Whereas the system shown in Figure B can confine charged particles (and has been used for a number of experiments), the special character of the Penning trap is given by the clever shaping and arrangement of the electrodes. As shown in Figure C, two end caps shaped as hyperbolae of revolution replace the parallel plates, and a ring-shaped center electrode defines the electrostatic potential on the edge of the trap.

The arrangement shown in Figure C not only leads to perfect confinement of individual charged particles but also allows the motion of a trapped particle to be separated into three independent harmonic motions. In order of decreasing frequency, the three motions are (1) the fast “cyclotron” motion of the charge around the magnetic field lines, (2) a slower oscillation in the direction of the magnetic field that is due to the electrostatic repulsion from the two end caps, and (3) a much slower drift motion that is due to the crossed electric and magnetic fields (see Figure D).

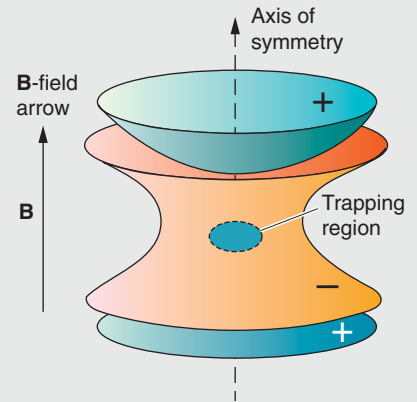


Figure C. The Penning Trap
The two endcaps, which are hyperbolae of revolution, replace the flat capacitor plates. The central ring electrode helps define the harmonic potential at the center of the trap.

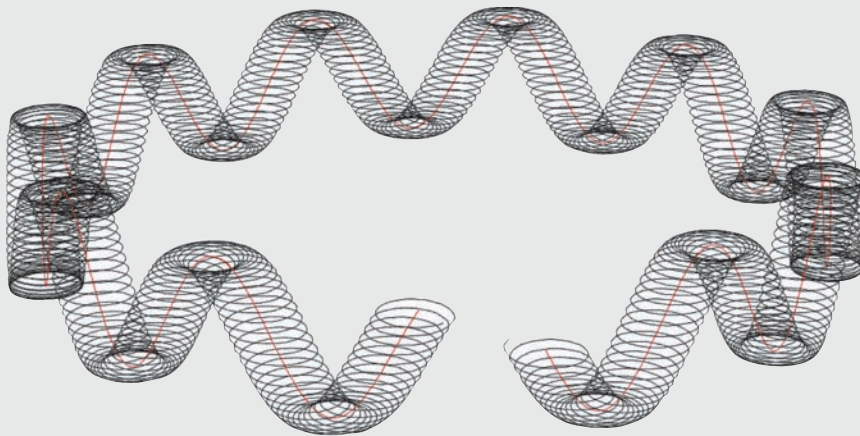


Figure D. Motion in the Penning Trap
The three-dimensional motion of an ion in the trap consists of three harmonic motions: a fast cyclotron motion, a slower up-down oscillation, and a slow circular drift motion.

The drift motion is easily understood if one focuses on the cyclotron motion. The positively charged particle is accelerated toward the negatively charged ring electrode as it moves away from the electrical center of the trap. This acceleration increases the radius of curvature for the outer half of the cyclotron motion. As the particle moves back toward the center during the second half of the cyclotron motion, it decelerates, and the radius of curvature decreases. The net effect is a distortion of the circular cyclotron motion into a spiral that bends around the electrical center of the trap, as seen in Figure E.

The harmonic motions account for the almost constant spacing between energy levels in Dehmelt’s geonium atom, but this orderliness is hardly noticeable in the roller-coaster-like motion of a trapped particle. If you actually want to experience the particle’s motion yourself, there is a carnival ride in which these three components of motion are present—but watch your stomach!

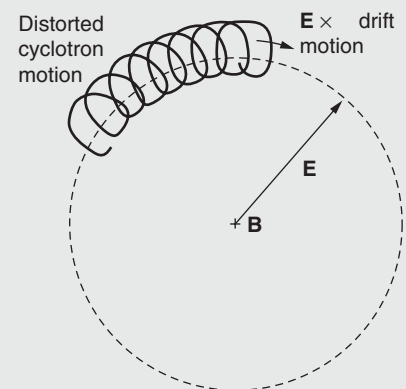


Figure E. Drift
Schematic of the drift motion that results from the crossed electric (E) and magnetic (B) fields.

Continued from page 265

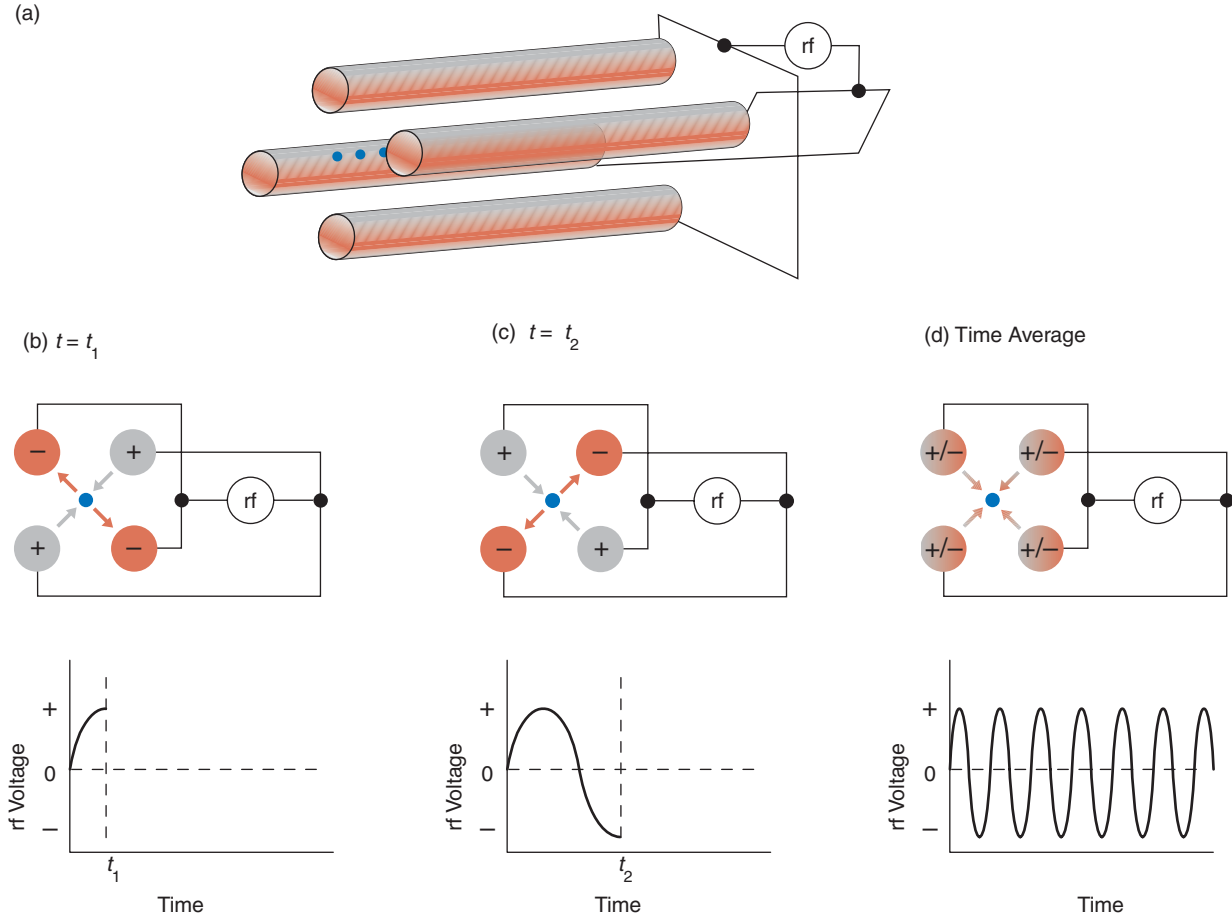


Figure 1. Principles of the Linear Paul Trap

(a) The linear Paul trap consists of four conducting rods. Two opposing rods are connected to one pole of a radio-frequency (rf) voltage source, whereas the remaining two are connected to the other pole. The axis of symmetry between the rods is the trap axis. (b) With the rods charged as shown, the resulting electric force pushes a positive ion to the negative rods and repels it from the positive ones. (c) Half an rf period later (see graph below), the polarity of all rods is reversed, and the direction of the force also reverses.

(d) If the polarity changes fast enough, a heavy ion becomes stuck in a rapid back-and-forth motion. Because the electric fields are at a minimum at the trap axis, an effective force pushes the ion toward the center, where it becomes trapped (although it is still free to move along the axis). The blue dots seen between the rods in part (a) represent a string of radially trapped ions. The string can be confined axially when a positively charged electrode (end cap) is placed at each end of the rods.

(relative to some arbitrary “zero” potential).¹ The positive ion feels a repulsive force from the positively charged conductors and is pushed toward the center of the trap. The ion simultaneously feels an attractive force due to the negatively charged

conductors and is pulled outwards.

If we now reverse the polarity of our four electrodes, interchanging plus for minus and minus for plus, the ion’s motion will begin to reverse. Where it was moving out, it will now be moving in, and vice versa. However, if the reversal takes place quickly, the “heavy” ion cannot easily respond; it has too much inertia to follow the fast changes in the electric field exactly.

Instead, the ion will respond to the time-averaged electric field. If we switch the polarity of the electrodes at a few megahertz (or a few million times a second) by applying a radio-frequency (rf) voltage and if the amplitude is correct, then the time-averaged field generates a harmonic pseudopotential with its minimum located at the trap axis. The ion is pushed to the bottom of the pseudopotential

¹ For reasons discussed on the previous two pages, all four conductors cannot be positively charged, or the electric fields within the trap would disappear.

well and becomes trapped forever—at least in principle—near the center.

To generate the mass selectivity sought by Paul, we add a positive direct-current (dc) component to the rf voltage. Positive ions outside a certain mass range feel less of a restoring force from the pseudopotential and are kicked out of the trap by the repulsive dc field. This technique is one of several that we can use to preferentially retain qubit ions instead of, say, a residual gas ion that may be present in the ultrahigh vacuum trap.

Of course, the ion's motion is still unrestricted in the direction parallel to the trap's axis. For confinement in this third dimension, we simply add an electric dc potential to a pair of "end electrodes" that are on either side of the region of interest. This axial field plugs up the escape route along the symmetry axis of the system, and the ion becomes trapped in three dimensions. By substantially reducing the ion's energy (cooling), we coax the ion into lying along the central portion of the trap axis, where the radial and axial confining potentials are at a minimum. If several very cold ions are in the trap, then they all fall to the center, and the mutual Coulomb repulsion between the ions causes them to line up neatly along the axis.

Motion in the Trap. Although the combination of rf and dc fields within the trap drives the ion into a complex radial motion, that motion is fully described by a set of differential equations called "Mathieu's Equations." The bound solutions of those equations yield a stability diagram that allows one to evaluate the effectiveness of the trap, given the values of several critical parameters, such as the amplitudes of the rf and dc components of the voltage, the rf, the ion mass, and the size of the trap (Dawson 1976).

As long as we keep the values of the critical operational parameters

within certain ranges, an ion will remain bound to the axis of the device. Furthermore, the magnitude of the restoring force of the pseudopotential holding the ions in the radial direction will remain directly proportional to the distance from the center—the hallmark of a harmonic potential.² In other words, to a good approximation, the ions will undergo harmonic oscillations in the radial direction with frequency ω_r (or with frequency ω_x and ω_y in case the potential is different in the x - and y -directions). This motion is commonly referred to as the secular motion.

The ions' motion can become distorted if the minima of the rf field and the pseudopotential are misaligned within the trap. Misalignment can occur because of some small asymmetry in the trap's construction or because of small dc patch potentials on the electrode surfaces. Regardless of the reason, misalignment will cause the ions to lie "off center" (that is, off the line where the rf field vanishes). Those ions will experience the strong gradient of the rf field and undergo rapid oscillations—at the frequency of the applied rf field—around their time-averaged equilibrium position. This so-called micromotion is the main source of ion heating. We can suppress the micromotion by adding a compensating dc voltage to some of the rf electrodes (or to auxiliary control electrodes) and thereby shift the ions' positions closer to the actual rf center.

In addition to exhibiting secular motion and the unwanted micromotion, an ion or, more important, a string of ions will also vibrate in the axial direction. The motion will be harmonic because the dc voltage on the end electrodes creates a harmonic

potential along some length of the trap axis. The vibrations are similar to those exhibited by a set of pendula connected to each other by springs; the swinging of one pendulum sets the others in motion (see Figure 2). Unlike vibrations of classical pendula, however, the vibrations exhibited by a string of ions are quantized; that is, the amplitude of the motion depends on the number of quanta (phonons) in the vibrational mode.

For N ions in a trap, there are N axial vibrational modes and an additional $2N$ modes for motions transverse to the axis. Each mode has a distinct vibrational frequency. The lowest-frequency (lowest-energy) vibration is the so-called common mode, in which the ions oscillate back and forth in unison along the axis. This mode figures heavily in the original quantum-computing scheme of Cirac and Zoller. Because all ions participate in the common-mode oscillation, when we add (or remove) a quantum of energy to this motion by interacting with one of the ions, we influence all other ions in the string. Any two qubits, regardless of the distance between them, can therefore be coupled together to perform logic operations.

Other proposals make use of some of the higher-frequency modes to couple qubits together (James 1998a). These modes have more-complex vibrational patterns and relatively higher excitation energies than the common mode, but it still takes very little energy to excite them. Even a string of very cold ions will vibrate in some intricate expression of the various modes, a problem that is addressed in the discussion of ion cooling.

If only a few ions are confined in the trap, the ions will align themselves linearly along the axis. But increasing the number of ions or increasing the dc voltage applied to the end electrodes introduces instabilities because

² To generate a pure harmonic potential, the four electrodes should have hyperbolic cross sections, but in practice we approximate that ideal shape with cylindrical rods.

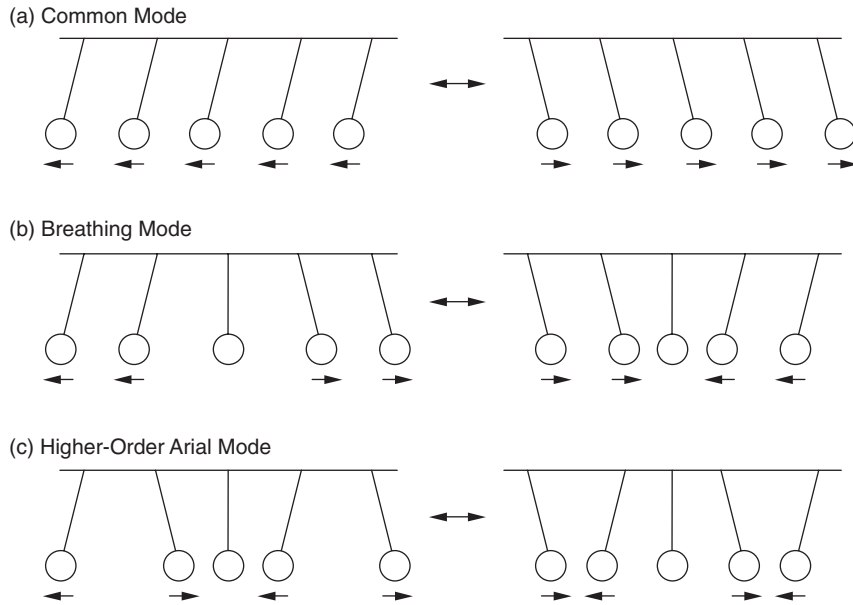


Figure 2. Vibrational Modes

A set of strongly coupled pendula can be used to envision the vibrational motion of a string of ions in a harmonic potential. These vibrational modes affect all ions simultaneously. If we set any one of the pendula in motion, the others will move. Similarly, if we grab hold of any pendulum and stop it, all others will stop. (a) The common mode (center-of-mass mode), in which all pendula swing one way and then the other, has the lowest frequency (lowest energy) of all modes. Using this mode to couple two qubits together in the trap is the basis of the Cirac-Zoller proposal. (b) The breathing mode, in which pendula at opposite ends move in opposite directions, has the next highest frequency. For an odd number of pendula, the middle one does not move. This mode is less susceptible to heating by external noise sources and has also been used to couple qubits. (c) Shown here is another higher-order mode. In an ion trap, the ions can vibrate in three dimensions; for N trapped ions, there are $3N$ vibrational modes.

the ions are effectively squeezed closer together. The Coulomb repulsion between neighboring ions becomes stronger than the radial restoring force, and the ions start buckling out into a zigzag pattern. When even more ions are added, the zigzag pattern develops into a complex three-dimensional helical structure (Raizen et al. 1992, Walther 1991, 1994). Some of the ions will move away from the axis and will experience the strongest micromotion heating—a situation clearly to be avoided. We have studied this transition from linear to three-dimensional structures in some detail (Enzer et al. 2000) and have quantified the param-

eter space available for quantum computing in a linear RFQ ion trap.

Elements of the Ion-Trap Quantum Computer

In 1994, inspired by the great success of ion traps in the field of precision measurements, Cirac and Zoller proposed that the RFQ ion trap had the right characteristics to support the long sequence of precision operations required for quantum computation. A string of ions trapped along the symmetry axis of the trap would be the quantum register of the computer. Each ion could be addressed by

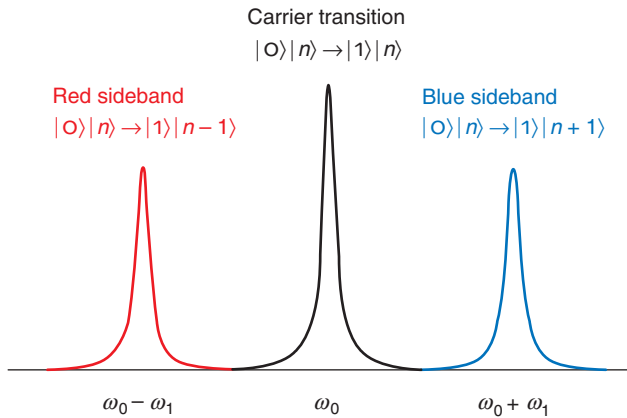
tightly focused laser beams, initialized to an arbitrary state, manipulated, and then probed at the end of a calculation. Most important, the isolation from the environment afforded by the trap would allow for long coherence times.

One- and Two-Qubit Operations.

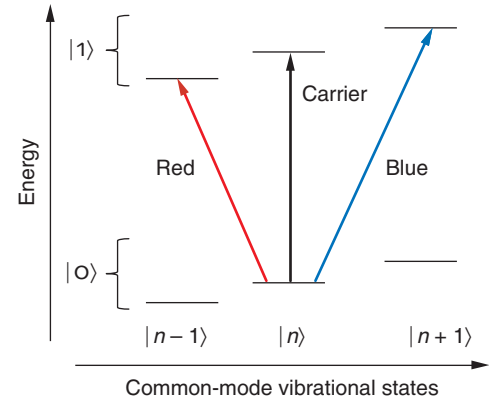
The logical qubit states $|0\rangle$ and $|1\rangle$ of the ion-trap quantum computer must be defined as they always are for any quantum computer. (To stress that the 0 and 1 used to designate the states are notational and have no numerical meaning, we have used a font different from the one for the numbers 0 and 1.) We simply identify the ion's electronic ground state with the qubit state $|0\rangle$ and a long-lived excited state with the qubit state $|1\rangle$.

We also want to apply a unitary transformation to a single qubit, that is, to implement a one-qubit gate, and rotate the qubit in Hilbert space to an arbitrary superposition of the $|0\rangle$ and $|1\rangle$ states. (Two-level systems and the rotation of a qubit in Hilbert space are discussed in the article "Quantum Information Processing" on page 2.) To do so, we subject the ion to a laser pulse of a specific amplitude, frequency, and duration. Assuming the ion is in its ground state, the laser pulse will cause the electron wave function of the target ion to evolve to some superposition of the ground and excited states. (We cause the electron to undergo part of a Rabi oscillation.) Illuminating the ion with a so-called π -pulse, for example, will evolve the electron wave function through half a Rabi oscillation period and leave the ion in the excited state. The qubit would have rotated from the $|0\rangle$ to the $|1\rangle$ state. If the duration of the pulse is halved (a so-called $\pi/2$ -pulse), the ion would be left in an equally weighted superposition of the ground and excited states. The qubit would have rotated to the $1/\sqrt{2}(|0\rangle + |1\rangle)$ state.

(a) Resolved Sideband Structure



(b) Resonant Transitions


Figure 3. Vibrational Sideband Spectrum

(a) An ion trap naturally couples an ion’s electronic excitations to its vibrational motion. Each electronic transition at resonant frequency ω_0 , known as the carrier frequency, is therefore accompanied by other sideband transitions. We show the two sidebands closest in frequency to the carrier: the lower-energy red sideband at frequency $(\omega_0 - \omega_1)$, and the higher-energy blue sideband at frequency $(\omega_0 + \omega_1)$. A laser with a sufficiently narrow linewidth can interact with the ion via a specific sideband or the carrier. (b) Interacting with a particular qubit (ion) via a sideband transition will change the qubit’s internal state and simultaneously the external state of all the qubits in the trap, either increasing the number of phonons in the common mode by one (excitation on the blue sideband) or decreasing the number by one (excitation on the red sideband).

While we can use laser pulses to interact with each qubit separately (and excite a qubit’s electronic, or internal, degrees of freedom), we can also use another laser to excite the trap’s vibrational modes and hence to interact with all qubits simultaneously. The latter process can be viewed as interacting with the qubits’ external degrees of freedom. The state of a string of j qubits in the trap is therefore explicitly given as

$$|q_1, q_2, \dots, q_j\rangle|n\rangle. \quad (1)$$

The first ket, $|q_1, q_2, \dots, q_j\rangle$, refers to the logical qubit states, with $q_j = 0$ or 1 . The second ket, $|n\rangle$, refers to the common-mode vibrational state, and the value of n , say, $0, 1, 2, \dots$, indicates the number of phonons in the common mode. (Although the string of qubits may initially be in another mode, we will restrict our attention to the common mode.) Thus, in the state

$$|q_1, q_2, \dots, q_j\rangle|0\rangle, \quad (2)$$

the ions are not vibrating because there are no phonons in the common mode. In the state

$$|q_1, q_2, \dots, q_j\rangle|1\rangle, \quad (3)$$

the common mode contains one phonon and all the ions are swaying in unison along the trap axis.

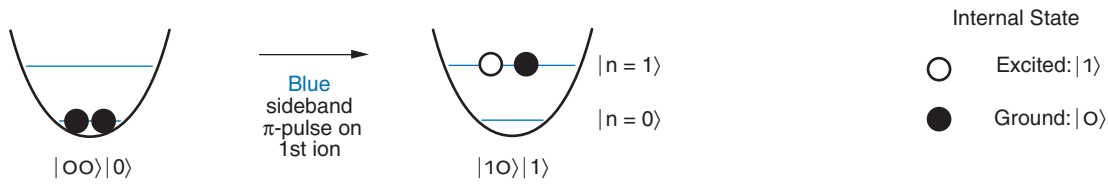
As mentioned earlier, the common mode is used as a “bus” that couples different ions together. To better understand this coupling, consider first that the frequency of the transition between the $|0\rangle$ state and the $|1\rangle$ state is ω_0 , and that the common-mode vibrational frequency ω_1 is much lower than ω_0 . Similar to the case of two coupled harmonic oscillators, the energy spectrum of the ion exhibits resonances not only at the “carrier” frequency ω_0 , but also at the “sideband” frequencies $\omega_0 \pm \omega_1$ (see Figure 3). The resonance with the higher frequency is commonly known as the blue sideband; the one with the lower frequency, as the red sideband. For cold ions, the linewidth $\Delta\omega_0$ of

the carrier transition is very narrow³ and is less than the energy difference between the carrier and either sideband. Thus, the sidebands and the carrier can be resolved within the cold ion’s frequency spectrum.

Now consider, for example, a procedure used to place two ions in an entangled state. Assuming that the ions are initially in the state $|\infty\rangle|0\rangle$, if we were to address the first ion with a π -pulse from a laser detuned to the blue sideband of the internal transition, both the internal and external states of that ion would be excited. Because an excitation in the common mode is felt by both ions, the result would be the two-qubit state $|10\rangle|1\rangle$. If, instead, we address the first qubit

³ The metastable excited state has a very long lifetime, which leads to a narrow linewidth according to Heisenberg’s uncertainty principle. To take an example from the Los Alamos experiment, a calcium ion excited to the $3^2D_{3/2}$ state will decay to the ground state only after an average delay of about 1 second, which results in a transition linewidth of about 1 hertz.

(a) Coupling Two Qubits through the Common Mode



(b) Entangling Two Qubits

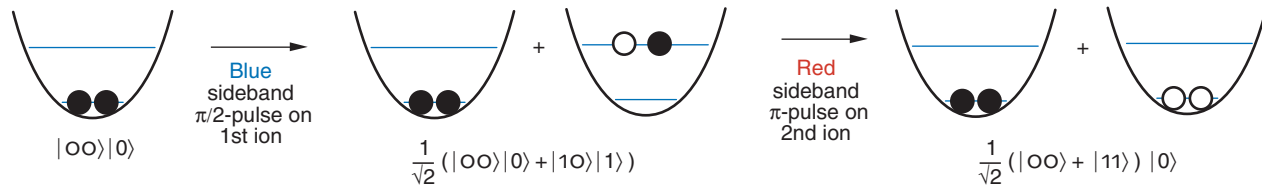


Figure 4. Using the Common Mode to Entangle Two Qubits

The vibrational state is indicated by the position of the ions on the rungs of a ladder in the harmonic potential well. In this diagram, the electronic ground state of an ion is indicated by a solid circle; the excited state, by an open circle. (a) Suppose two qubits are initialized to the state $|00\rangle|0\rangle$. Addressing the first qubit with a π -pulse from a laser tuned to the blue sideband will excite the ions to the state $|10\rangle|1\rangle$. The first ion is in its electronic excited state, while the second remains in its electronic ground state. Because the common mode affects all ions, both ions are excited to the $|n = 1\rangle$ vibrational state. (b) Two qubits can be

entangled if we illuminate the first qubit with a $\pi/2$ -pulse on the blue sideband. The ions are placed in a superposition of states: $1/\sqrt{2}(|00\rangle|0\rangle + |10\rangle|1\rangle)$. If the second ion is then illuminated by a π -pulse from a laser tuned to the red sideband, it can absorb the photon only if energy is available from the vibrational mode. Thus, ion 2 is excited only if ion 1 was excited; it remains in the ground state if ion 1 was in the ground state. The two-ion system therefore exhibits the strong correlation of observables, which according to Bohr, define the condition of entanglement. The end result of the operation is the entangled state $1/\sqrt{2}(|00\rangle + |11\rangle)|0\rangle$.

with a $\pi/2$ -pulse (see Figure 4), both qubits are placed in a superposition of the two states, namely,

$$1/\sqrt{2}(|00\rangle|0\rangle + |10\rangle|1\rangle) . \quad (4)$$

We then address the second ion (which is still in its ground state) with a π -pulse tuned to the red sideband. The laser energy is too low to excite directly the ground-to-excited-state electronic transition, but the transition still occurs if extra energy can be taken from the common mode. The end result is that all phonons have been removed from the quantum register at the end of the operation, and we create a two-qubit entangled state:

$$1/\sqrt{2}(|00\rangle + |11\rangle)|0\rangle . \quad (5)$$

We can no longer describe the system as individual ions being in the ground or the excited state. The result of a measurement on one ion is strongly correlated to the status of the other ion. Notice that this procedure works equally well if one or more ions are placed in between the first and second ions because the excitation of the common mode is shared by all ions.

Besides defining the individual operations just described, Cirac and Zoller also spelled out in detail the steps needed to perform a “controlled-not” (**cnot**) gate. In such an operation, a “target qubit” flips its state only if a second qubit, the “control qubit,” is originally set to its logical $|1\rangle$ value. Dave Wineland’s group at NIST first implemented the **cnot** gate in an ion trap in 1995 (Monroe et al. 1995),

albeit using only a single ion in the trap. (The two states of the control qubit were the two lowest-energy trap vibrational states.) Still, because any computation can be performed with a number of two-qubit **cnot** gates, together with some single-qubit gate operations, the realization of a **cnot** gate in a quantum computer is an important benchmark.

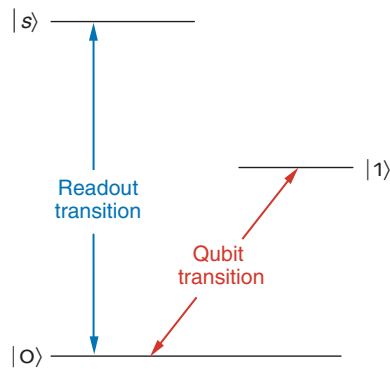
Readout. At the end of any quantum calculation, the individual qubits in the quantum register will be in defined states, which must be read out with high fidelity. A powerful readout tool makes use of the phenomenon of quantum jumps (Sauter et al. 1986, Bergquist et al. 1986). The readout method is easily understood when one examines the generic ion-level scheme

shown in Figure 5. The ion has two states that we identify as the logical qubit states $|0\rangle$ and $|1\rangle$. But the ions used for quantum computation also have a short-lived excited state $|s\rangle$ that is accessible from one of the qubit states, the $|0\rangle$ state, by laser excitation. When the laser drives the $|0\rangle \rightarrow |s\rangle$ transition for a long period, the ion fluoresces and emits a huge number of photons. If that transition is not accessible because the ion was in the $|1\rangle$ state, there will be no fluorescence. Detection of a fluorescence signal, therefore, tells us that the qubit is in the $|0\rangle$ state, and we observe the “jumps” of the ion between the $|0\rangle$ and the $|1\rangle$ state as distinct jumps in the intensity of the fluorescence.

This type of readout will destroy any quantum information contained in the qubit state and will yield a purely For example, suppose the ion is in an equal superposition of the states $|0\rangle$ and $|1\rangle$; then probing the ion once with the laser will not reveal the original state of the qubit. If we want to get a reading on the ratio of the two different states in a superposition, we will have to repeat the measurement multiple times and resort to a statistical description.

If we want to maintain the quantum character of the ion’s state at the end of a particular calculation, we may resort to a different scheme. Consider an ion placed in a high-quality optical cavity, which is tuned to the resonance of the internal transition in the ion. If the ion is in the state $|1\rangle$, a photon is emitted into the cavity; if it is in the state $|0\rangle$, no photon is emitted. If the ion is in a superposition state, the photon field in the cavity will end up in a superposition between the states consisting of one photon in the cavity and no photon in the cavity. Thus, the quantum state of an ion or an atom can be transferred to a photon (Parkins and Kimble 1999, Mundt et al. 2002). This state could be transferred through optical fibers to a different trap and then

(a) Generic Three-Level System



(b)

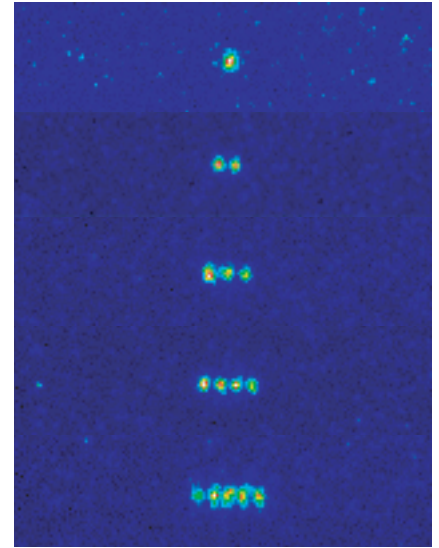


Figure 5. Readout Using Quantum Jumps

(a) A generic three-level scheme for ions in a trap is illustrated. The qubit states $|0\rangle$ and $|1\rangle$ are typically the ground state and a long-lived excited state, respectively. The state $|s\rangle$ is short lived and is coupled to the ground state. If the ion is in the ground state, a laser can drive the $|0\rangle \rightarrow |s\rangle$ transition many times per second, and the ion will fluoresce. If the ion “jumps” to the $|1\rangle$ state, there will be no fluorescence, so that the presence or absence of a large fluorescence signal reveals the state of the qubit. (Alternatively, one can use two long-lived ground-state hyperfine levels as qubits and construct a similar readout scheme.) (b) This composite image shows strings of calcium ions that were laser-cooled to near rest in the Los Alamos quantum computation ion trap. The spacing between the ions is approximately $30\ \mu\text{m}$. About 10^8 photons are absorbed and reemitted each second during the time the readout laser is irradiating the ion. That photon flux is easily detectable with a charge-coupled device (CCD) camera. The fluorescence is actually bright enough to be seen with the naked eye, except that for calcium, the readout transition is at $397\ \text{nm}$ and is outside the range of sensitivity for the human eye.

transferred into another ion—and so, the quantum Internet is born!

The Los Alamos Ion-Trap Quantum Computing Experiment

Currently, every implementation of ion-trap quantum computing uses qubits that are composed of two long-lived internal states of the trapped ions (the ground state and a metastable excited state, or two hyperfine sublevels of the ground state) and has the qubits communicating with each other through the trap’s vibrational modes. Many different ion

species can serve as qubits, and numerous qubit schemes are possible. While the previous section discussed ion-trap quantum computers in general terms, this section describes an experiment developed at Los Alamos, in which calcium ions were used.

We initially chose to use calcium for a number of reasons, including the following: All the wavelengths needed for cooling and manipulation are, at least in principle, accessible by relatively inexpensive diode lasers; the lifetime of the metastable state allows a reasonable number of coherent operations to be performed; and the calcium isotope of interest is most abundant and can easily be loaded

into the trap. But the basic quantum computational schemes outlined earlier can be implemented with any element that displays an ionic-level structure similar to that of calcium, such as the other alkali-earth elements beryllium, magnesium, strontium, and barium. At this stage of experimentation, all alkali-earth ions are essentially interchangeable, and for mostly technical reasons, calcium has recently been replaced with strontium in our quantum computing experiment. (Some of that work is described in the article, “Quantum Information with Trapped Strontium Ions” on page 178.) In addition, other ions, such as mercury and ytterbium, also exhibit level schemes that are applicable to quantum computation, albeit with slightly different technical approaches. As ion-trap quantum computers become more sophisticated, the choice of ion species will become a larger issue.

Our trap is a linear Paul trap, about 1 centimeter in length and 1.7 millimeters in diameter, with a cylindrical geometry, as seen in Figure 6(a). We create the strong, radial confinement fields by applying a few hundred volts of rf amplitude at approximately 8 megahertz to two opposing rods. The remaining two rods are rf-grounded. The axial confinement, which prevents the ion from leaking out of the trap along the symmetry axis, is produced by a direct current of about 10 volts applied to each of the conical end caps. This combination of the rf and dc fields leads to an axial oscillation frequency ω_1 for the common mode of a few hundred kilohertz and a radial oscillation frequency of $\omega_r \approx 1$ megahertz.

Additional dc potentials can be applied to four support rods, which are not shown in Figure 6(a) but are located outside the actual trap electrodes. In this way, one can center the ion string on the electrical symmetry axis of the trap and thus mini-

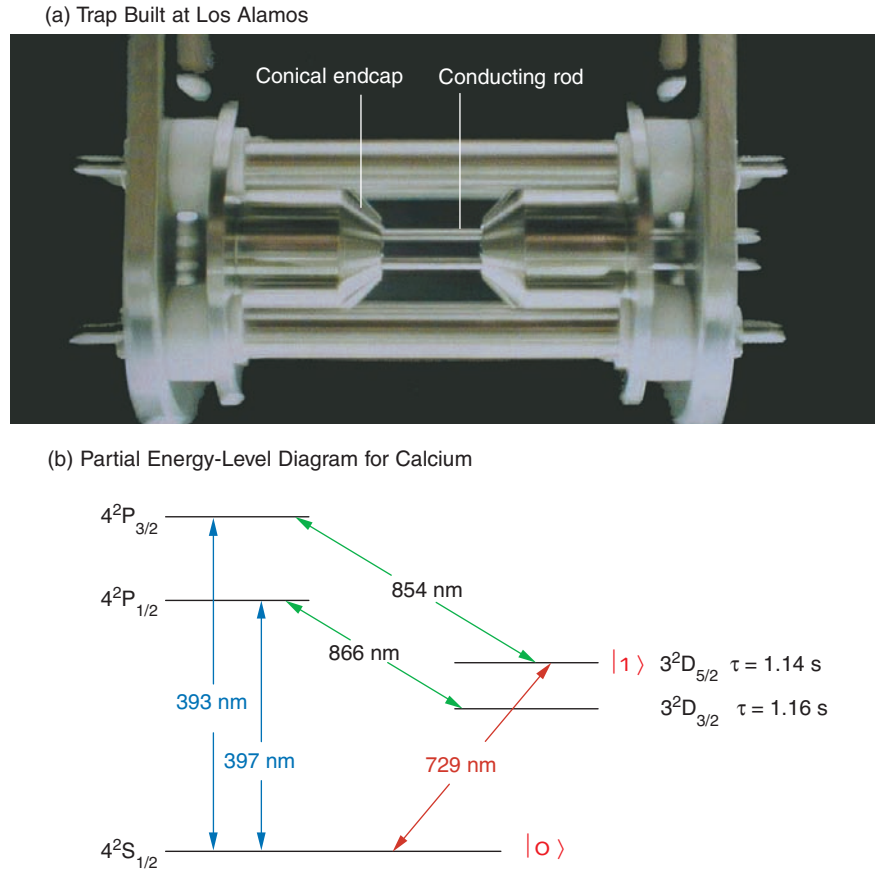


Figure 6. The Los Alamos Linear Paul Trap

(a) The trap built at Los Alamos for quantum computation is about 1 cm in length and 1.7 mm in diameter. An electric field of a few hundred volts oscillating at 8 MHz is applied to two of the conducting rods. The other two rods are RF grounded. About 10 V of a direct current is placed on the conical end caps. (b) This illustration shows a partial energy-level diagram for calcium (not to scale) and shows the wavelengths of some transitions relevant to our quantum computing scheme. The qubit transition is shown in red.

mize the amount of heating produced by micromotion.

Figure 6(b) shows a schematic diagram of the internal-level structure of calcium ions and gives the wavelengths of the relevant transitions. (Any other alkali-like ion would have a similar structure.) The $4^2S_{1/2}$ ground state and the metastable $3^2D_{5/2}$ excited state are used to form the logical qubit states $|0\rangle$ and $|1\rangle$, respectively. The metastable excited state has a lifetime of about 1 second, which is long enough to allow an interesting number of computational steps to be performed before decoherence (resulting

from spontaneous emission from the excited state) can destroy the internal state of the quantum register.

To load the ions into the trap, we cross a beam of calcium atoms that is produced by heating a small calcium-filled reservoir with a beam of electrons emitted by an “electron gun.” (The electron gun is essentially identical to the one inside a computer monitor or a television screen.) These two beams are aligned so that they cross each other within the effective volume of the trap, that is, within the cylindrical volume that fits between the four rods and the two end caps. The atoms

that are ionized by electron impact suddenly feel the confining forces of the electric fields and become trapped.

Cirac and Zoller (1995), as well as other authors, proposed initializing the computer in the state

$$|00 \dots 0\rangle|0\rangle ; \quad (6)$$

that is, all qubits are in their electronic and vibrational ground states.

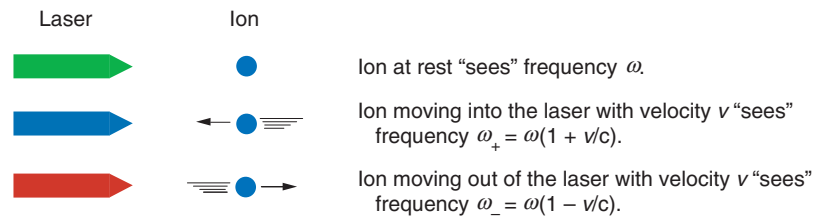
However, the temperature⁴ of the newly trapped ions is very high, since their energy is given by a combination of the temperature of the calcium oven and the energy imparted to the ion by the electric field. (The latter energy varies, depending on where the ionization occurs.) In order to reach the initial state and then to perform quantum logic operations, the ions' temperature must be reduced to its lowest possible value. Cooling the ions takes place in two steps described in the next two sections.

Doppler Cooling of Calcium Ions.

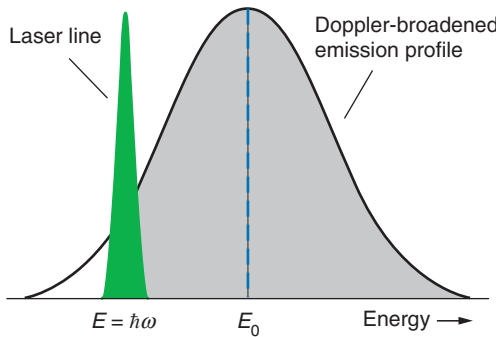
As its name suggests, this first cooling step makes use of the Doppler effect, whereby the relative motion between a source and an observer causes a change in the observed frequency of an acoustic or electromagnetic wave. For example, the sound of a siren on an ambulance or a police car changes its pitch depending on whether the vehicle moves toward you or away from you. Similarly, an ion or atom will absorb or emit photons of different frequencies (energies), depending on its motion relative to the light source. Although an ion in the trap is localized by electric fields and its average velocity is zero, the variation of the instantaneous velocity, as the ion jiggles back and forth due to thermal motion, causes the inherent emission and/or

⁴ We often refer to ion temperature rather than energy because the ions show a distribution of energies over time.

(a) Frequency Shifts Due to the Doppler Effect



(b) Detuning



(c) Doppler Cooling of Calcium

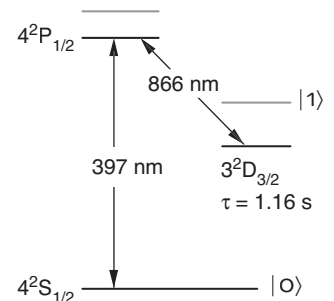


Figure 7. Doppler Cooling of Ions

(a) When interacting with a laser of frequency ω , an ion at rest sees the native laser frequency. If the ion is moving, this frequency is shifted by the Doppler effect. An ion moving into the laser beam “sees” the laser frequency Doppler-shifted toward a higher frequency, ω_+ , while the ion moving in the direction of the laser beam “sees” the frequency ω_- . (b) This first-order Doppler effect is eliminated in ion traps because the average velocity is zero. However, because of its thermal motion, the ion has a probability to absorb photons at any frequency within its Doppler-broadened absorption profile. Similarly, it has a probability to emit a photon over a range of frequencies within its emission profile. Suppose the laser is tuned below the ion’s resonance frequency ω_0 so that $\omega < \omega_0$. When the ion moves into the laser beam, it will absorb a photon because it sees the laser frequency Doppler-shifted close to its resonance frequency ($\omega_+ \sim \omega_0$). The ion absorbs a laser photon of energy $E = \hbar\omega < \hbar\omega_0$, but on average it reemits a photon with higher energy (from the gray region). Because it loses energy during each cycle of absorption and emission, the ion cools rapidly to the limit of this method, which is imposed by the recoil energy the ion experiences upon reemission of the photon. For typical parameters of our trap, calcium will reach a vibrational level of approximately $|n = 10\rangle$ to $|n = 30\rangle$ at the end of the Doppler cooling. (c) The transitions used to Doppler-cool calcium ions are shown.

absorption profile of an ionic transition to become much broader than the natural linewidth of the transition (second-order Doppler broadening). For “hot” ions, the Doppler-broadened linewidth is typically much greater than the laser linewidth.

To implement Doppler cooling, we tune a laser to a frequency below the resonance frequency of a transition in the ion (Figure 7). Only when it is moving at a certain velocity toward

the laser can the ion absorb these “off-resonance” photons, because only then does it “see” the laser frequency shift into resonance. However, as a result of its random jiggling, the ion has a probability to emit photons at any frequency within its Doppler-broadened emission line profile. One can easily see from Figure 7 that the ion has a greater probability to emit a photon with a higher frequency than the absorbed photon. On average,

more energy is emitted than absorbed, which leads to a cooling of the ion.

For rapid cooling, a large number of photons must be absorbed and emitted, and therefore Doppler cooling is performed on a transition that can be cycled rapidly. We use the 397-nanometer transition from the $4^2S_{1/2}$ to the $4^2P_{1/2}$ state. The lifetime of the $4^2P_{1/2}$ state is about 10 nanoseconds, so the ion can absorb and reemit about 10^8 photons per second. Unfortunately, the $4^2P_{1/2}$ state has a chance of roughly 1 in 15 to decay into the metastable $3^2D_{3/2}$ state, which has a lifetime of about 1 second. The ion then takes so long to return to the ground state that it would be lost from the cooling cycle. To avoid this outcome and force ions to return from the $D_{3/2}$ level to the cooling cycle, we irradiate the ion with two lasers: the cooling laser at 397 nanometers and a “repump” laser at 866 nanometers.

Doppler cooling has its limits. Conservation of momentum guarantees that, after emitting a photon in one direction, the ion recoils in the opposite direction. Although this recoil energy is small, eventually it counteracts any cooling effects. For calcium ions, the Doppler limit is equivalent to a temperature of about 3 microkelvins. At that temperature, the kinetic energy of the ions is significantly less than the mutual Coulomb repulsion between ions. Essentially, they do not have enough kinetic energy to leap-frog each other, so the cold ions remain frozen in their relative locations and form a string. The photos in Figure 4 are examples of ion strings that were realized in our trap. At a 200-kilohertz common-mode frequency, the spacing between ions is about 30 micrometers.

Even at a temperature of 3 microkelvins, however, the ions have enough energy to occupy any of several vibrational modes, with many phonons per mode. (The specific dis-

tribution of states depends on the ions’ temperature and the frequency of each mode.) Here, we will restrict our attention to the common mode. After Doppler cooling, the ions in the trap can typically occupy the states from $|n = 10\rangle$ to about $|n = 30\rangle$. Getting the qubits into the common-mode ground state ($|n = 0\rangle$), therefore, requires an additional cooling scheme.

Sideband Cooling of Calcium.

We recall that ions in the trap can couple their internal degrees of freedom with their external motion, which leads to sidebands at $\omega_0 \pm \omega_1$, where ω_1 is the common-mode frequency, in the absorption spectrum. For cold ions with a minimal Doppler linewidth, these sidebands are resolvable from the carrier—see Figure 8(b). Thus, an ion can absorb photons not only at the carrier frequency ω_0 of their internal $|0\rangle \rightarrow |1\rangle$ transition but also on the upper and lower sidebands at the frequencies $\omega_0 \pm \omega_1$.

Assuming all ions are in the state $|0\rangle|n\rangle$, we can tune a laser with a suitably narrow linewidth to the red sideband—photon energy $[E_- = \hbar(\omega_0 - \omega_1)]$ —and excite one of the ions to the state $|1\rangle|n - 1\rangle$. In essence, energy is removed from the vibrational mode (the occupation number is reduced by one phonon) and is used to make up the deficit in photon energy. After its radiative lifetime, the ion can decay to one of three states: the state $|0\rangle|n - 2\rangle$, by emitting a photon with energy $[E_+ = \hbar(\omega_0 + \omega_1)]$; the state $|0\rangle|n - 1\rangle$, by emitting a photon with energy $[E_0 = \hbar\omega_0]$; or a return to its initial state, by emitting a photon with energy $[E_- = \hbar(\omega_0 - \omega_1)]$ —see Figure 8(c). On average, the ion loses one vibrational photon of energy $E = \hbar\omega_1$ for each excitation–decay cycle. Because we started somewhere around $|n = 30\rangle$, we need about 30 cycles to bring the vibrational mode to its ground state (provided there are no competing effects that heat the ions while they are being cooled).

Unfortunately, the long lifetime of the $3^2D_{5/2}$ state is now a hindrance. In principle, we can scatter only one photon per second using this transition, which would render the process of cooling from $|n = 30\rangle$ to $|n = 1\rangle$ unacceptably slow. Heating processes—micromotion heating or simply radiative heating from other noise sources in the system—are much faster, and we would be unable to reach the desired starting point of all qubits being in the internal and external ground states.

To speed things up, we artificially shorten the lifetime of the $3^2D_{5/2}$ state by introducing an alternate decay route via the $4^2P_{3/2}$ state (Marzoli et al. 1994). We irradiate the ion not only with a laser tuned to 729 nanometers (to drive the S–D transition), but also with a second laser tuned to 854 nanometers—see Figure 8(d). The second laser pumps the ion from the D- to the P-state, from which the ion rapidly returns to the ground state. By carefully choosing the amplitude of the 854-nanometer laser, we can design the effective lifetime of the $3^2D_{5/2}$ state according to our needs, and our calcium ion can jump down the ladder of harmonic-oscillator levels in just 3 to 30 milliseconds.

In a real system, the cooling power from the lasers will always be in competition with external heating processes. Although no clear theoretical explanation of these processes has emerged, many possibilities have been discussed in the literature, and the relevant scaling laws with trap parameters have been developed (James 1998b). Typical candidates—besides micromotion heating, which can be avoided by carefully tuning the trap voltages—are fluctuating contact potentials on the trap electrodes (originating from insulating deposits on the electrodes), which have a frequency component at the trap’s resonant frequency.

In the absence of a proper theoretical description of ion heating, we can

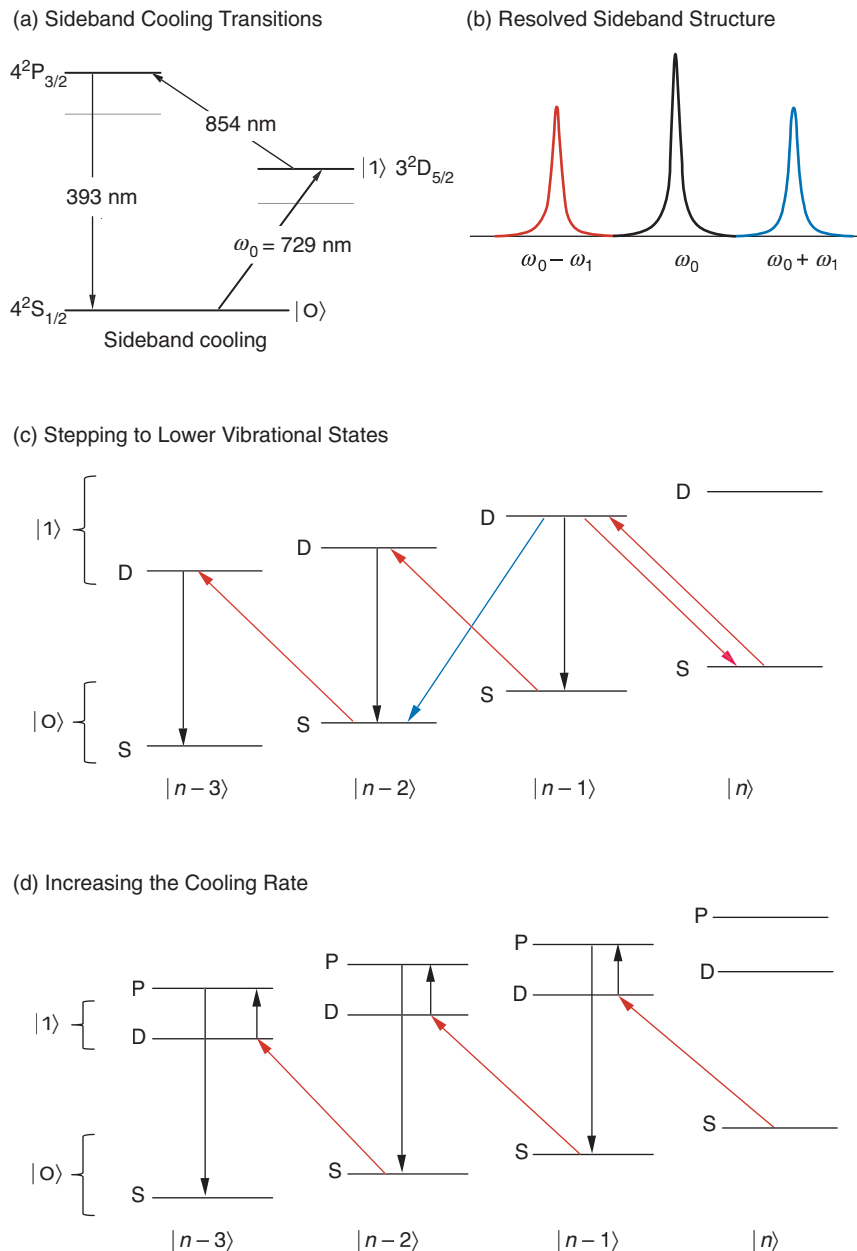


Figure 8. Sideband Cooling

(a) This partial energy-level diagram shows the transitions we use for sideband cooling of calcium ions. (b) When the linewidth of the carrier transition (frequency ω_0) is very narrow and the Doppler broadening is minimal, the ion's vibrational sidebands can be resolved. (c) The figure shows several vibrational levels for the $|0\rangle \rightarrow |1\rangle$ carrier transition. If a single ion is initially in the state $|0\rangle|n\rangle$, then illuminating the ion with a laser tuned to the red sideband will excite the ion to the state $|1\rangle|n-1\rangle$. The latter state will decay to $|0\rangle|n-2\rangle$ or $|0\rangle|n-1\rangle$, or it will go back to $|0\rangle|n\rangle$. On average, the number of phonons in the mode decreases by 1 after each excitation/emission. (d) The lifetime of the upper level may be artificially shortened if that level is coupled to an auxiliary one with a higher decay rate. The faster decay will increase the cooling rate.

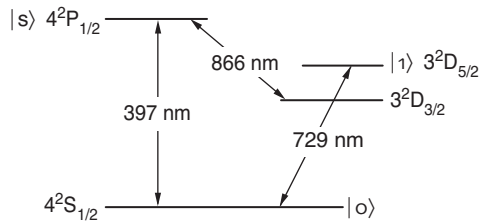
turn to empirical data accumulated from a number of different experiments. The adopted procedure is to cool the ions to as low a temperature as possible and then turn off the lasers responsible for the cooling. After a variable delay time, we measure the ion's temperature using sideband spectroscopy. Quentin Turchette and coworkers (2000) conducted the most complete study of this type when they looked at heating effects in traps of different sizes. The separate traps had also undergone different preparation "rituals." The studies suggest a strong dependence on trap size, that is, on the distance between the ions and the trap electrodes. When the studies are combined with observations made by Rainer Blatt's group at the University of Innsbruck, one is led to believe that "bigger is better." But Ralph deVoe with IBM has recently reported that hardly any heating was observed over a short period in a miniaturized trap.

Clearly, we have much to learn before we can understand the heating of ions in rf traps. The comforting thought is that, in all cases, the time scale for heating from the ion's ground state can be kept long, compared with the time required for a reasonable number of quantum manipulations. Furthermore, heating times are typically longer than times for other decoherence processes.

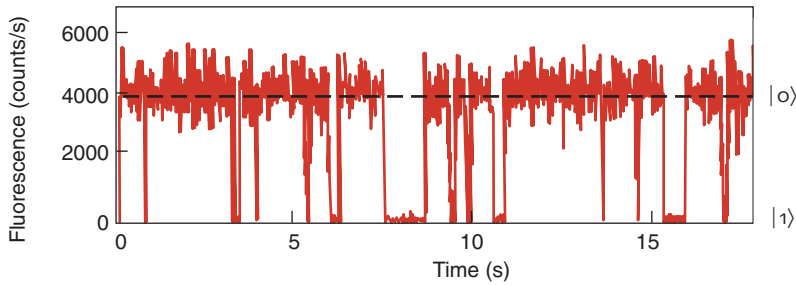
Readout of the Calcium Ion.

We use the $|s\rangle = 4^2P_{1/2}$ excited state in calcium for readout, the same state that is used for Doppler cooling. As discussed earlier, this state has a lifetime of only 10 nanoseconds and is accessed from the ground state by a laser tuned to 397 nanometers. An ion in the $|0\rangle$ state will absorb and reemit about 10^8 photons per second when the laser drives the $|0\rangle \rightarrow |s\rangle$ transition. (Because the $4^2P_{1/2}$ state can also decay to the long-lived $3^2D_{3/2}$ state, we simultaneously irradiate the ion with a laser tuned to

(a) Calcium Readout Transitions



(b) Quantum Jumps, Single Ion



(c) Quantum Jumps, Two Ions

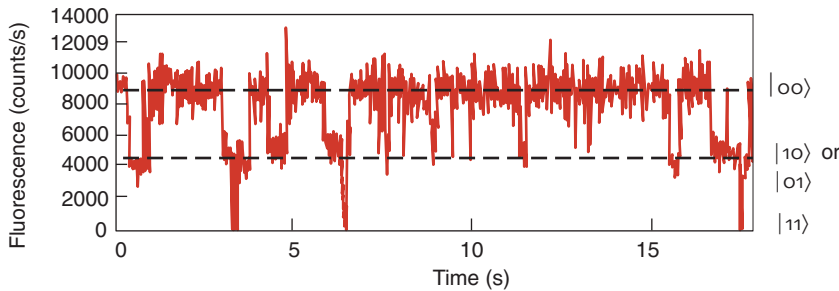


Figure 9. Readout of Qubits

(a) Shown here are the readout transitions for calcium. (b) For this readout experiment, a single ion interacts with two lasers: a low-intensity laser that drives the qubit transition $|0\rangle \rightarrow |1\rangle$ and a second laser that drives the readout transition $|o\rangle \rightarrow |s\rangle$. The fluorescence signal from that transition, nominally around 4,000 counts per second, is recorded with a simple rate meter. When the qubit is in the $|o\rangle$ state, we can drive the readout transition, but if the ion occupies the state $|1\rangle$, the fluorescence disappears. We can distinguish between the $|o\rangle$ and $|1\rangle$ states with nearly 100% fidelity. (c) The state of two ions can also be distinguished. No count corresponds to the state $|1, 1\rangle$; 8000 to 10,000 counts per second correspond to the state $|o, o\rangle$; 4000 counts per second, to either $|o1\rangle$ or $|10\rangle$. (In the last case, our experimental setup does not allow us to distinguish between the two states.)

866 nanometers to return the ion to the $4^2P_{1/2}$ state.) Even with a modest photon-collection efficiency of about 10^{-4} , which is due to experimental considerations (we cannot bring a lens too close to the ions without blocking access to the trap), we can easily detect the photons scattering from the ion with a charge-coupled device (CCD) camera.

In Figure 9, we show a sample trace

of the detected photon counts for a single ion in the trap. The fluorescence signal is nominally about 10^4 counts per second. We randomly excite the ion with a laser tuned to 729 nanometers, and each time it “jumps” from the $|0\rangle$ state to the $|1\rangle$ state, the signal disappears. Figure 9 also shows the fluorescence from a set of two ions. The different levels of intensity are for both ions being excited (no fluorescence),

for one of the two ions being excited (intermediate fluorescence), and finally, for both ions being in the ground state (full fluorescence). Although it is easy to distinguish among these cases, determining which of the two ions is in the ground state for the intermediate fluorescence level is difficult. We must look at the ions individually, by focusing the laser on one ion at a time, and then convert to the single-ion measurement.

Ferdinand Schmidt-Kaler and his colleagues from the Innsbruck group have used this readout technique with three ions, which were spaced at about 6 micrometers from each other in the trap. They cooled the ions to the $|000\rangle|n\rangle$ state, and all three were emitting photons on the readout transition. The scientists then pointed a sharply focused laser at 729 nanometers onto one of the ions and placed it in the $|1\rangle$ state (the dark state). The measured crosstalk among neighboring ions was less than 1 percent, so the state of the chosen qubit could be determined with about 99 percent fidelity (Nägerl et al. 1999).

Important Developments

A *Popular Mechanics* article from 1949 stated, “Where a calculator on the ENIAC (electronic numerical integrator and calculator) is equipped with 18,000 vacuum tubes and weighs 30 tons, computers in the future may have only 1000 tubes and weigh only one and a half tons.” That observation did not turn out to be entirely correct. How could anyone have foreseen the development of transistors and integrated solid-state circuitry or the remarkable parallel developments that have culminated in today’s supercomputers?

We are still in the “vacuum-tube” era of quantum computation, and if asked two years ago about the future of ion-trap-based quantum computers,

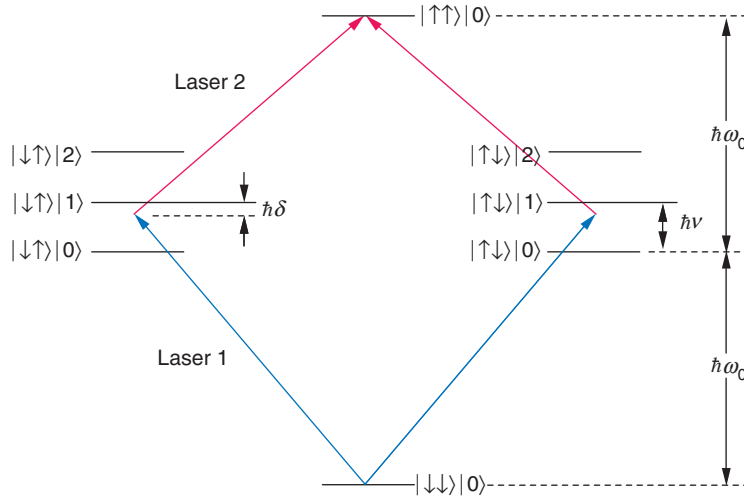


Figure 10. Four-Particle Entanglement

The figure shows the relevant energy levels and transition frequencies used to create deterministic multiparticle entanglement. A two-ion scheme is illustrated. The $|\uparrow\uparrow\rangle|0\rangle$ excited state has an energy of $2E_0 = 2\hbar\omega_0$. The $|\uparrow\downarrow\rangle|1\rangle$ and $|\downarrow\uparrow\rangle|1\rangle$ excited states, in which the internal state of one of the ions is excited and both ions go into a vibrational excited state, has an energy $E_1 = \hbar(\omega_0 + \nu)$. Lasers tuned to energies $E_1 + \delta$ and $E_1 - \delta$, where δ is a predetermined laser detuning, can directly excite the ions to the $|\uparrow\uparrow\rangle|0\rangle$ state. Pulsing the two lasers for a time $t = \pi/(2\Omega)$, where Ω is an effective Rabi frequency, will place the ions in the entangled state $|\Psi_2\rangle = 1/\sqrt{2}(|\uparrow\uparrow\rangle - i|\downarrow\downarrow\rangle)$. The scheme can be generalized to any number of ions and has been used to create entangled states of up to four ions.

(Figure reproduced with permission from *Nature*.)

I would have been hesitant to promise much. I may have argued that the systems we were looking at were mere demonstrations, designed to help us understand the fundamental physics issues behind qubits and that the prospects for scaling these devices up to a larger number of qubits were doubtful. Even today I could argue that, while the computing scheme of Cirac and Zoller is in principle scalable (Hughes et al. 1996), it has yet to be realized with two qubits.

However, because much has happened in the ensuing two years, included here are descriptions of just a few of the many important developments that have put the ion-trap quantum computer back on the track for scalable technologies. Similar to the transition from vacuum tubes to solid-state devices (even if not quite as fundamental), these developments do not invalidate any of the previous

achievements and underlying principles but are unpredicted and significant enhancements of available technology.

Four-State Entanglement. To take full advantage of the power of quantum computation, we need to generate entanglement between an arbitrary number of qubits. But generating any entangled state is difficult. In the case of photons, entanglement is achieved by means of a statistical process. Many pairs of photons are created by a method known as parametric down-conversion, whereby a high-energy photon, after entering a special type of crystal, has a certain probability to convert into two photons, each with half the initial energy. In a few cases, two photons emerge in an entangled state. (See the article “Quantum State Entanglement” on page 52.) We can typically produce about 1000 entan-

gled pairs per second, but if we look for entanglement of three or even four photons, the likelihood of finding such a state becomes unacceptably small for practical purposes (30 per second for 3 photons and a few per year for four photons).

Thus, quantum computing took a leap forward when the NIST team in Boulder demonstrated that it could produce an entangled state of up to four ions “on demand” (Sackett et al. 2000). Based on a proposal by Anders Mølmer and Klaus Sørensen (1999) from the University of Aarhus in Denmark, the NIST team around Chris Monroe and David Wineland demonstrated the feasibility of entanglement of two and four ions in a deterministic way. With a single-pulse operation of two lasers, the desired state could be produced with a high degree of certainty.

To understand the technique, consider two spin-half particles confined in a harmonic well and coupled by vibrational degrees of freedom. (The spin description is equivalent to our previous picture of two internal states in an ion.) The NIST group used the two ground-state hyperfine levels of $^9\text{Be}^+$ ions as an effective spin-half system, with $|\downarrow\rangle = |F = 2, m_F = -2\rangle$ and $|\uparrow\rangle = |F = 1, m_F = -1\rangle$. The energy levels of the system are shown in Figure 10, where $\hbar\omega_0$ is the internal energy splitting of each particle and ν is the oscillation frequency of the particular collective mode of the particles in the trap.

The group used standard laser-cooling and optical-pumping techniques to prepare the particles in their spin-down internal state and in the ground state of their collective motion: $|\Psi\rangle = |\downarrow\downarrow\rangle|0\rangle$. Laser pulses at $\omega_0 + (\nu - \delta)$ and $\omega_0 - (\nu - \delta)$, where δ is the detuning from the resonance (refer to Figure 10), then drive the two-step transition from $|\downarrow\downarrow\rangle|0\rangle$ to $|\uparrow\uparrow\rangle|0\rangle$. If the detuning δ is sufficiently large, the intermediate states $|\uparrow\downarrow\rangle|1\rangle$

and $|\downarrow\uparrow\rangle|1\rangle$ are negligibly occupied, and no motional excitation occurs in the process. Applying the laser fields for a time $t = \pi/(2\Omega)$, where Ω is the Rabi oscillation frequency for the transition, results in the final wave function

$$|\Psi_2\rangle = 1/\sqrt{2} (|\uparrow\uparrow\rangle - i|\downarrow\downarrow\rangle) , \quad (7)$$

which is the desired maximally entangled state.

It turns out that this process is entirely scalable for an even number of N qubits and can generate the N -particle entangled state

$$|\Psi_N\rangle = 1/\sqrt{2} \times (|\uparrow\uparrow\dots\uparrow\rangle - i^{N+1}|\downarrow\downarrow\dots\downarrow\rangle) . \quad (8)$$

If N is an odd number, the state $|\Psi_N\rangle$ can still be produced, provided one rotates each qubit independently. The NIST scientists have used this method with two and four ions in the trap, but they also caution that the experimentally realized states $|\Psi_2\rangle$ and $|\Psi_4\rangle$ are not fully entangled. Each state shows some degree of decoherence. Although the amount of decoherence in $|\Psi_4\rangle$ was more than what could be tolerated for quantum computing, the achievement of reliably creating a four-particle entangled state on demand cannot be underestimated.

In a later development, the NIST group showed that the maximally entangled states of a pair of trapped ${}^9\text{Be}^+$ ions could be used as a decoherence-free subspace for protecting one qubit of information against dephasing errors (Kielinski et al. 2001). The decoherence-free subspace, also called a noiseless subsystem, is spanned by the two orthogonal states

$$|\Psi_+\rangle = 1/\sqrt{2} (|\downarrow\uparrow\rangle + i|\uparrow\downarrow\rangle) , \text{ and} \\ |\Psi_-\rangle = 1/\sqrt{2} (|\downarrow\uparrow\rangle - i|\uparrow\downarrow\rangle) . \quad (9)$$

These states serve as the logical qubit for storing information. It is easy to

see that all superpositions of these maximally entangled states are invariant under transformations that apply the phase change $|\uparrow\rangle \rightarrow e^{i\zeta}|\uparrow\rangle$ simultaneously to both ions. This so-called collective dephasing is thought to be a major source of decoherence for trapped ions.

In the NIST experiment, an arbitrary state of one qubit was encoded in the decoherence-free subspace of two ions:

$$\alpha|\uparrow\rangle + \beta|\downarrow\rangle \rightarrow \alpha|\Psi_+\rangle + \beta|\Psi_-\rangle . \quad (10)$$

The encoded information was subjected to engineered dephasing errors or ambient errors, and then the encoding procedure was reversed to recover the original information. The data showed unequivocally that the noiseless subsystem protects the information from collective dephasing errors for a time up to ten times longer than the typical decoherence time and that collective dephasing is indeed a major source of errors in ion traps. One could imagine that this type of robust storage might enable the operation of a quantum computer constructed from an array of ion traps as opposed to a single trap. (For an introduction to the theory of noiseless subsystems, see the article “Introduction to Quantum Error Correction” on page 188. A nuclear magnetic resonance experiment demonstrating noiseless subsystems is presented in the article “Realizing a Noiseless Subsystem in an NMR Quantum Information Processor” on page 260.)

Broadband Cooling. The second important recent result is the selective enhancement of the probability of cooling ions by electromagnetically induced transparency (EIT). The scheme of Cirac and Zoller has the qubits coupled together by means of the common vibrational mode, in which all ions oscillate back and forth in unison along the trap axis. However, even two trapped ions have

an extra degree of freedom in the axial motion, namely, the breathing mode, in which ions on opposite sides of the string move 180° out of phase (refer to Figure 2). Each additional ion opens up three more vibrational modes to the ion string. Every mode of frequency ν can be assigned an average quantum number n_ν .

The initial scheme of Cirac and Zoller requires a mode to have $n_\nu = 0$ in order to be used for computational operations. For small numbers of ions, we reach this state by the standard sideband-cooling methods discussed earlier. As seen in Figure 11(a), the ion has a number of transition possibilities: Excitation on the lower sideband will cool the ion, excitation on the upper sideband will cause heating, and transitions on the carrier will cause diffusion. In sideband cooling, we use an ultranarrow laser and excite only the lower sideband so that $|n\rangle \rightarrow |n-1\rangle$.

For a large number of qubits, however, the sheer number of higher modes makes it technically difficult, if not impossible, to use standard sideband-cooling methods. Not only would we have to identify and excite the lower-sideband transitions for each and every mode, but the spectrum becomes so “dense” that the upper sidebands of a neighboring internal transition can overlap the lower sidebands of another. Cooling one mode could actually heat another. Furthermore, the “overhead” needed to control and cool these modes is daunting: large numbers of laser pulses, constant retuning of the lasers from one mode to the next, and tight control of the qubit register throughout the cooling stage.

For efficient (and simultaneous) cooling of more than one mode, broadband cooling would be required, even though that would seemingly exacerbate the problem of unwanted excitation. But recent work by Blatt’s group at the University of Innsbruck

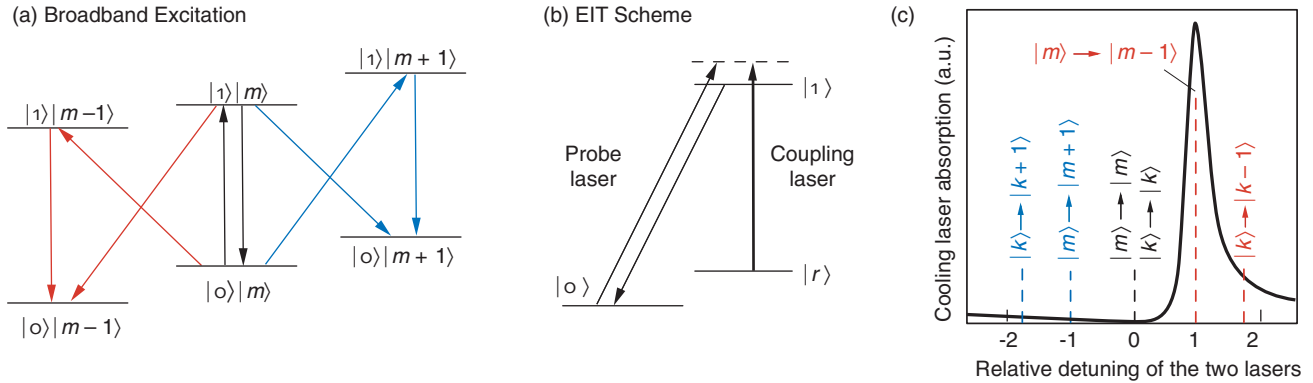


Figure 11. EIT Cooling

Sideband cooling of a multi-ion string that is accessing many excited vibrational modes is very difficult in that the sideband structure becomes dense and complicated. EIT cooling permits broadband cooling of several vibrational modes, $|m\rangle$, $|k\rangle$, ..., simultaneously. (a) When a broadband probe laser is applied to the $|o\rangle|m\rangle \rightarrow |1\rangle|m\rangle$ transition, both cooling (red) and heating (blue) transitions can occur. (b) When a second coupling laser excites the $|r\rangle \rightarrow |1\rangle$ transition, the ion's absorption profile becomes modified. Proper choice of laser detuning (to the dashed state) suppresses heating transitions.

This result is evident in figure (c), where the solid line gives the absorption profile for the EIT scheme. For proper tuning of the lasers, the absorption strength for the transition $|m\rangle \rightarrow |m\rangle$ is zero and a strong asymmetry between $|m\rangle \rightarrow |m+1\rangle$ and $|m\rangle \rightarrow |m-1\rangle$ transitions is introduced. This asymmetry in absorption between the blue and the red sideband also holds for higher-frequency vibrational modes ($|k\rangle \rightarrow |k\pm 1\rangle$), allowing simultaneous cooling of several different modes with one broadband laser. [Figure was adapted from Schmidt-Kaler (2001) with permission from the authors.]

may make broadband cooling possible (Morigi et al. 2000, Roos et al. 2000). The group adopted the EIT technique to selectively enhance the probability of exciting cooling transitions rather than heating transitions in the ion.

The necessary asymmetry between lower and higher sidebands can be achieved as follows: Consider a three-level system with two lower levels and one shared excited state—see Figure 11(b). Using a strong coupling laser between one of the ground states and the upper state creates so-called light shifts (that is, shifted energy levels, as seen by another probe laser). For a detuning of the coupling laser above the resonance, a probe laser sees an absorption profile that shows zero absorption for a detuning equal to the coupling laser, the so-called Fano profile—see Figure 11(c). Therefore, such a probe would be transparent for that exact detuning—the EIT phenomenon. In order to obtain optimum cooling using these EIT resonances, the detunings are chosen such that the carrier transition is exactly located at the EIT resonance

(that is, it is not excited at all because of that quantum interference), and the maximum absorption is chosen to be around the lower sideband frequency.

Because the absorption profile generated in this manner is fairly wide (much wider than the natural width of the transition used for traditional sideband cooling), the asymmetry between heating and cooling transitions exists for many modes. Several different modes can be cooled simultaneously with a single operation. This technique reduces the overhead for laser-cooling of multi-ion strings and also eases the requirements for laser stability, which are very strict for standard sideband cooling.

To show that EIT cooling can simultaneously cool vibrational modes with significantly different frequencies of oscillation, the Innsbruck group chose to cool the axial mode and the radial mode of a single ion confined in a three-dimensional Paul trap at 3.3 megahertz and 1.6 megahertz, respectively (Schmidt-Kaler et al. 2001). In a linear trap, the nearby

modes (“spectator” modes) are not used for the computation directly; they are coupled to and may affect the common mode. The group achieved ground-state populations of 73 percent for the axial and 58 percent for the radial mode. Although this result is certainly not as satisfactory as that achieved by sideband cooling (because of the smaller absorption asymmetries), it is certainly sufficient for cooling (and thus suppressing) those modes. The EIT method promises the possibility of cooling all spectator modes of a multiqubit quantum register with a single operation. That would allow the more elaborate (individual) sideband cooling scheme to be used on only the mode needed for calculations.

Outlook

Many systems have been proposed in the last several years as potential candidates for becoming quantum computers, including laser-cooled trapped ions (Cirac and Zoller 1995),

nuclear magnetic resonance (Gershenfeld and Chuang 1997, Cory et al. 1998), cavity quantum electrodynamics (Ye et al. 1999), and more recently, superconducting devices, quantum dots, and silicon-based solid-state devices.

From the preliminary experiments performed by several groups worldwide, it is apparent that the existing ion traps are adequate to hold and manipulate small numbers of qubits. Although five to ten qubits hardly a computer make, these numbers are large enough to make the technology well worth pursuing. Ion traps will be a potent tool for exploring, for example, the possibility of creating entangled states of large numbers of qubits. Investigations of the type described here will help us identify the relevant physics issues that must be addressed to achieve computational gains.

We should also expect that many of the technologies now being pursued for quantum computation will be superseded by even more promising ideas. One such idea is to scale up to a larger number of qubits by multiplexing several ion traps with a specific trap that contains a few qubits acting as the central processor. After implementing part of a quantum algorithm, the qubits would be shuffled into one of several storage traps, thus allowing new qubits to be loaded into the processor. Recent work also suggests that we could transfer the internal quantum states of a string of ions in a trap to a set of photons in a high-finesse cavity. The quantum information could then be transferred through optical fibers into a second cavity and fed back into an ion string in a different trap. Developments like this will surely continue to happen and will allow us to explore quantum computation well beyond the current state of the art.

As we get closer to realizing a small quantum processor, the “time scales” of a particular system become more relevant. In general, the hierar-

chy of time scales present in an ion-trap quantum computer is very promising. Manipulations on quantum registers can be done in microseconds, while disturbances by the environment have been shown to be avoidable for milliseconds. The inherent decoherence time of the quantum state is longer still, for it is limited by the lifetime of the upper qubit state, which is about 1 second in calcium. The decoherence time can be increased even more by an appropriate choice of ions (for example, ytterbium) or by stable ground-state hyperfine levels used as logical qubit states.

It is important to point out that despite the revolutionary advances in computers during the last 50 years, the fundamental principle of computation has not changed. Today’s fastest supercomputer operates according to the same rules as the ENIAC.

Quantum computation, however, represents a paradigm shift in information processing. Although a future quantum computer may not look anything like our current ion trap, the experience and knowledge we gain now will be of fundamental importance to our understanding this new paradigm of computing.

For some researchers, building a quantum computer to break secure codes is an important, and certainly challenging, goal. But for me and most of my colleagues, performing experiments that Erwin Schrödinger and Albert Einstein only dreamed of and thus gaining a deep understanding of this “inconceivable” quantum world are far larger rewards. Perhaps we will encounter some failure of conventional quantum mechanics, or perhaps the problems of decoherence will forever keep the quantum realm out of our classical grasp. In any event, the future will be exciting for both quantum physics and computation. ■

Acknowledgments

The work described in this article is the result of a close and fruitful collaboration among numerous experimental scientists over several years. I wish to thank them all for their help in the laboratory, as well as for the many intellectual discussions that helped me understand in depth the fundamental principles involved. Over the years, I enjoyed working with Daphna Enzer, John Gomez, Mark Gulley, Paul Kwiat, Steve Lamoreaux, Glen Peterson, Vern Sandberg, Martin Schauer, Dale Tupa, and Justin Torgerson on developing capabilities for quantum computation with trapped ions. In addition, I had the privilege to travel to many groups pursuing this dream and learned much of what I know today about the ion-trap quantum processor from many interactions with my friends and colleagues Rainer Blatt, Ferdinand Schmidt-Kaler, Dietrich Leibfried, Christoph Nägerl, and many others at the University of Innsbruck in Austria; with Andrew Steane, David Lucas, and Derek Stacey from the Clarendon Laboratory in Oxford, the United Kingdom; and last but not least, with all the members in Dave Wineland’s group at NIST, in Boulder (too numerous to name here individually), who took me on challenging excursions into the quantum world of trapped ions, as well as into the all too classical world of mountain biking.

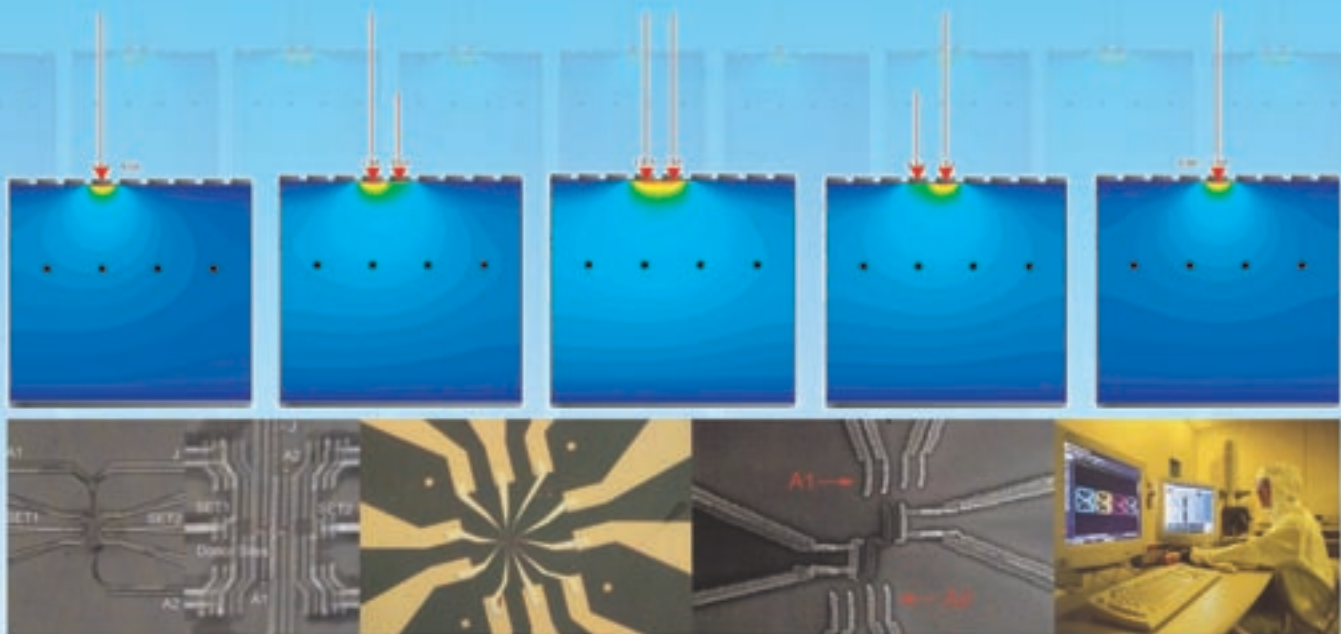
Further Reading

- Bergquist, J. C., R. G. Hulet, W. M. Itano, and D. J. Wineland. 1986. Observation of Quantum Jumps in a Single Atom. *Phys. Rev. Lett.* **57**: 1699.
- Cirac, J. I., and P. Zoller. 1995. Quantum Computations with Cold Trapped Ions. *Phys. Rev. Lett.* **74**: 4091.
- Cory, D. G., M. D. Price, and T. F. Havel. 1998. Nuclear Magnetic Resonance Spectroscopy: An Experimentally Accessible Paradigm for Quantum Computing. *Physica D* **120** (1–2): 82.

- Dawson, P. H., ed. 1976. *Quadrupole Mass Spectrometry and Its Applications*. Chaps. II and III. Amsterdam: Elsevier Scientific Publishing Co.
- Dehmelt, H. G. 1967. Radiofrequency Spectroscopy of Stored Ions, Part I. *Adv. At. Mol. Phys.* **3**: 53.
- . 1969. Radiofrequency Spectroscopy of Stored Ions, Part II. *Adv. At. Mol. Phys.* **5**: 109.
- . 1981. Coherent Spectroscopy on Single Atomic System at Rest in Free Space 2. *J. Phys. (Paris)* **42**: 299.
- . 1988. A Single Atomic Particle Forever Floating at Rest in Free Space: New Value for Electron Radius. *Physica Scripta*. **T22**: 102.
- Enzer, D. G., M. M. Schauer, J. J. Gomez, M. S. Gulley, M. H. Holzschleiter, P. G. Kwiat et al. 2000. Observation of Power-Law Scaling for Phase Transitions in Linear Trapped Ion Crystals. *Phys. Rev. Lett.* **85** (12): 2466.
- Gershenfeld, N. A., and I. L. Chuang. 1997. Bulk Spin-Resonance Quantum Computation. *Science* **275**: 350.
- Huang, X.-P., J. J. Bollinger, T. B. Mitchell, and W. M. Itano. 1998. Phase-Locked Rotation of Crystallized Non-Neutral Plasmas by Rotating Electric Fields. *Phys. Rev. Lett.* **80**: 73.
- Hughes, R. J., D. F. V. James, E. H. Knill, R. Laflamme, and G. F. Petschek. 1996. Decoherence Bounds on a Quantum Computation with Trapped Ions. *Phys. Rev. Lett.* **77**: 3240.
- Hughes, R. J., D. F. V. James, J. J. Gomez, M. S. Gulley, M. H. Holzschleiter, P. G. Kwiat, et al. 1998. The Los Alamos Trapped Ion Quantum Computer Experiment. *Fortschr. Phys.* **46** (4–5): 329.
- James, D. F. V. 1998a. Quantum Dynamics of Cold Trapped Ions with Application to Quantum Computation. *Appl. Phys. B* **66** (2): 181.
- . 1998b. Theory of Heating of the Quantum Ground State of Trapped Ions. *Phys. Rev. Lett.* **81**: 317.
- Kielpinski, D., V. Meyer, M. A. Rowe, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland. 2001. A Decoherence-Free Quantum Memory Using Trapped Ions. *Science* **291**: 1013.
- Marzoli, I., J. I. Cirac, R. Blatt, and P. Zoller. 1994. Laser Cooling of Trapped Three-Level Ions: Designing Two-Level Systems for Sideband Cooling. *Phys. Rev. A* **49**: 2771.
- Mølmer, K., and A. Sørensen. 1999. Multiparticle Entanglement of Hot Trapped Ions. *Phys. Rev. Lett.* **82** (9): 1835.
- Monroe, C., D. M. Meekhof, B. E. King, W. M. Itano, and D. J. Wineland. 1995. Demonstration of a Fundamental Quantum Logic Gate. *Phys. Rev. Lett.* **75**: 4714.
- Morigi, G., J. Eschner, and C. H. Keitel. 2000. Ground State Laser Cooling Using Electromagnetically Induced Transparency. *Phys. Rev. Lett.* **85** (21): 4458.
- Mundt, A. B., A. Kreuter, C. Becher, D. Leibfried, J. Eschner, F. Schmidt-Kaler, and R. Blatt. 2002. Coupling a Single Atomic Quantum Bit to a High Finesse Optical Cavity. <http://eprints.lanl.gov/quant-ph/0202112>.
- Nägerl, H. C., D. Leibfried, H. Rhode, G. Thalhammer, J. Eschnerr, F. Schmidt-Kalerr, and R. Blatt. 1999. Laser Addressing of Individual Ions in a Linear Ion Trap. *Phys. Rev. A* **60** (1): 145.
- Neuhauser, W., M. Hohenstatt, P. E. Toschek, and H. Dehmelt. 1980. Localized Visible Ba⁺ Mono-Ion Oscillator. *Phys. Rev. A* **22**: 1137.
- Parkins, A. S., and H. J. Kimble. 1999. Quantum State Transfer between Motion and Light. *J. Opt. B, Quantum Semiclassical Opt.* **1** (4): 496.
- Paul, W., H. P. Reinhard, U. von Zahn. 1958. Electrical Mass Filters as Mass Spectrometers and Isotope Filters. *Z. Phys.* **152**: 143.
- Raizen, M. G., J. M. Gilligan, J. C. Bergquist, W. M. Itano, and D. J. Wineland. 1992. Ionic Crystals in a Linear Paul Trap. *Phys. Rev. A* **45**: 6493.
- Roos, C. F., D. Leibfried, A. Mundt, F. Schmidt-Kaler, J. Eschner, and R. Blatt. 2000. Experimental Demonstration of Ground State Laser Cooling with Electromagnetically Induced Transparency. *Phys. Rev. Lett.* **85** (26): 5547.
- Sackett, C. A., D. Kielpinski, B. E. King, C. Langer, V. Meyer, C. J. Myatt, et al. 2000. Experimental Entanglement of Four Particles. *Nature* **404**: 256.
- Sauter, Th., W. Neuhauser, R. Blatt, and P. E. Toschek. 1986. Observation of Quantum Jumps. *Phys. Rev. Lett.* **57**: 1696.
- Schmidt-Kaler, F., J. Eschner, G. Morigi, C. F. Roos, D. Leibfried, A. Mundt, and R. Blatt. 2001. Laser Cooling with Electromagnetically Induced Transparency: Application to Trapped Samples of Ions or Neutral Atoms. *Appl. Phys. B* **73** (8): 807.
- Steane, A., C. F. Roos, D. Stevens, A. Mundt, D. Leibfried, F. Schmidt-Kaler, and R. Blatt. 2000. Speed of Ion-Trap Quantum-Information Processors. *Phys. Rev. A* **62**: 042305.
- Tanoudji, C. C., B. Diu, and F. Laloe. 1977. *Quantum Mechanics*, Vol. I, Chap. IV. B. 3a, p. 403. New York: John Wiley and Sons.
- Turchette, Q. A., D. Kielpinski, B. E. King, D. Leibfried, D. M. Meekhof, C. J. Myatt, et al. 2000. Heating of Trapped Ions from the Quantum Ground State. *Phys. Rev. A* **61**: 063418.
- Walther, H. 1991. In *Light Induced Kinetic Effects on Atoms, Ions, and Molecules*, p. 261. Edited by L. Moi et al. (Pisa, Italy: ETS Editrice).
- . 1994. Atoms in Cavities and Traps. *Adv. At. Mol. Opt. Phys.* **32**: 379.
- Wineland, D., P. Ekstrom, and H. Dehmelt. 1973. Monoelectron Oscillator. *Phys. Rev. Lett.* **31**: 1279.
- Ye, J., D. W. Vernooy, and H. J. Kimble. 1999. Trapping of Single Atoms in Cavity QED. *Phys. Rev. Lett.* **83** (24): 4987.

Michael Holzschleiter received his M.S. and Ph.D. degrees in physics from the University of Mainz in Germany, where he studied electrons trapped in Penning traps. As a postdoctoral researcher at Texas A&M University, Michael used trapped ions to study collisional processes of astrophysical relevance. Later, as an assistant professor at Texas A&M, he participated in a collaboration with Los Alamos National Laboratory on trapping antiprotons. In 1986, he joined Los Alamos and pioneered the dynamic trapping of high-energy particles in Penning traps. He became principal investigator of the Los Alamos antiproton experiment, which was installed at CERN in 1992. On the basis of his technique's success, he formed an international collaboration, ATHENA, to create antihydrogen atoms at rest for ultrahigh precision studies of the symmetries between matter and antimatter and served as spokesman for this collaboration from 1995 through 1999. Back in Los Alamos, Michael applied his expertise in ion traps toward a Los Alamos trapped-ion quantum computer, for which he designed the radio-frequency quadrupole trap and helped build and operate the first-generation experiment until the fall of 2001. His research interest continues in the application of the ion-trap technology to a wide variety of physics problems.





Toward a Silicon-Based Nuclear-Spin Quantum Computer

Developing the technology for a scalable, solid-state quantum computer

*Robert G. Clark, P. Chris Hammel, Andrew Dzurak, Alexander Hamilton, Lloyd Hollenberg, David Jamieson, and Christopher Pakes
as told to Jay Schecker*

One of the major challenges in quantum computing is to identify a system that can be scaled up to the number of qubits needed to execute nontrivial quantum algorithms. Peter Shor's algorithm for finding the prime factors of numbers used in public-encryption systems (numbers that typically consist of more than a hundred digits) would likely require a quantum computer with several thousand qubits. Depending on the error correction scheme appropriate to the particular computer, the required number could be much larger. Although ion-trap or nuclear-magnetic-resonance (NMR) quantum "computers" containing a few (less than 10) qubits have been successfully

demonstrated, it is not clear that those systems can be scaled up.

Solid-state quantum computers may be more likely candidates. As a result, a number of solid-state schemes have been proposed, most of which fall into two categories: The physical qubits are spin states of individual electrons or nuclei, or they are charge or phase states of superconducting structures.

A particularly promising scheme is the silicon-based nuclear-spin computer, proposed a few years ago by Bruce Kane (1998), then of the University of New South Wales in Sydney, Australia, and now of the University of Maryland in College Park, Maryland. Shown in Figure 1,

the Kane computer architecture consists of a linear array of phosphorus atoms embedded beneath the surface of a pure silicon wafer. Each phosphorus atom serves as a qubit, and the linear array forms the computer's quantum register.¹ The entire wafer is placed in a strong, static magnetic field \mathbf{B}_0 at sub-Kelvin temperatures, and alignment of the phosphorus atom's nuclear spin as parallel or antiparallel to \mathbf{B}_0 corresponds to the $|0\rangle$ and $|1\rangle$ states of the qubit, respectively. (Throughout this article, we will follow the convention of Kane

¹ Two-dimensional qubit arrays are also possible but require complex electrode geometries and additional insulators.

and use the notation $|0\rangle$, $|1\rangle$ to designate both qubit and nuclear spin states. We will use arrows, $|\downarrow\rangle$ or $|\uparrow\rangle$, to designate electron spin states.)

In order to execute a quantum algorithm, we need to rotate individual qubits in Hilbert space and couple two qubits together. We accomplish both operations with an array of gate electrodes² that lies on top of the wafer but is isolated from the pure silicon by a thin insulating layer of silicon dioxide (SiO_2). Referring to Figure 1, notice that each A-gate sits precisely above a phosphorous atom and each J-gate lies between adjacent atoms. As discussed later, a small positive voltage applied to the A-gate gives independent control of the qubit directly under the gate, while voltage applied to the J-gate allows coupling two qubits together through an electron-mediated interaction.

The Center for Quantum Computer Technology (CQCT), headquartered in Sydney, Australia, and Los Alamos National Laboratory are trying to fabricate Kane's silicon-based quantum computer. Although we can call upon the technology, techniques, and collective experience of the huge silicon semiconductor industry, building the computer is still a daunting enterprise. We need to manipulate individual phosphorus atoms and place them precisely within a defect-free, isotopically pure silicon matrix. We need to create metal gates on the nanoscale, lay them within a few atoms of each other, and then ensure that each gate is aligned properly over the buried qubits. Simply put, the ability to do this level of nanofabrication does not exist at this time.

Employing a “see-what-works-best” strategy, we have initiated parallel research approaches for nearly every fabrication stage. If one approach fails, we still have a backup. Our current focus is on developing a prototype

² In this context, a “gate,” like a transistor gate, is a device used for controlling charge motion. It does not refer to a logical operation such as a **not** gate.

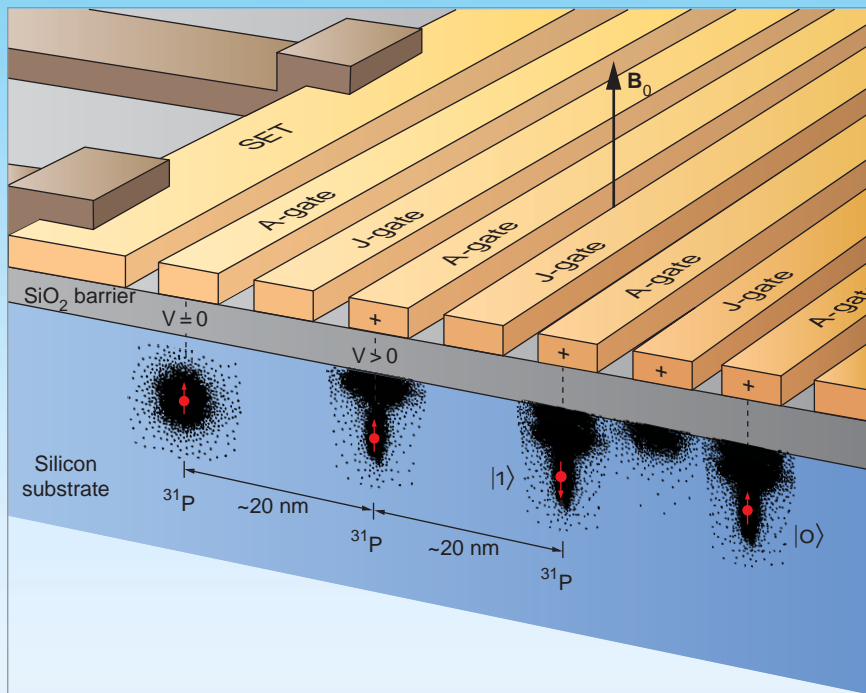


Figure 1. Schematic of the Silicon-Based Quantum Computer

The architecture of Kane's solid-state quantum computer has a linear array of phosphorous donor atoms buried into a pure silicon wafer, with an interdonor spacing of about 20 nm. In the presence of a large magnetic field and at sub-Kelvin temperatures, the nuclear spins of the donor atoms can be aligned either parallel or antiparallel with the field, corresponding to the $|0\rangle$ and $|1\rangle$ qubit states, respectively. An array of metal gates lies on the surface of the wafer, isolated from the silicon by a barrier layer of SiO_2 . The A-gates lie directly above the donor atoms and enable individual control of single qubits. The J-gates lie between the donors and regulate an electron-mediated coupling between adjacent nuclear spins, thus allowing for two-qubit operations. Readout of the qubit is performed with either a single electron transistor, as shown, or with a magnetic-resonance force microscope (MRFM, not shown). The electron clouds shown here are schematic—the actual situation is more complicated.

two-qubit device. By stretching many technologies beyond their heretofore-assumed limits, we have come tantalizingly close to achieving that goal. In the sections that follow, we describe the computer and some critical technologies in greater detail, and we also outline our progress in building a prototype.

Design Features of the Silicon-Based Computer

In our solid-state architecture, individual phosphorus atoms are embedded in a sea of silicon. These

two elements were chosen for several reasons, the first and foremost being that phosphorous is the standard dopant for conventional silicon-based semiconductor devices and there is tremendous working knowledge of phosphorus and silicon from the conventional computing industry.

The second reason stems from the need to control the spin environment. We require our qubits to have nuclear spin $I = 1/2$, but we also want the surrounding environment to be spin free. Otherwise, unwanted spin-spin interactions between the qubit and any nearby nuclear spins would compromise the coherent states needed for

quantum computation. The only stable phosphorous isotope, phosphorous-31, is a spin-1/2 nucleus, so nature has automatically satisfied our qubit criterion. Creating a spin-free environment, however, is a little more difficult.

Natural silicon contains a mixture of three isotopes: silicon-28, -29, and -30. Whereas the even-numbered isotopes are spin free, silicon-29 has a spin of $I = 1/2$. As a result, we estimate that to do quantum computation, we will need to reduce the silicon-29 content in our silicon wafer to about one part in 10^5 . Those stringent isotopic purity levels can be reached with current technology.

Finally, the nuclear spin of a phosphorous atom in a silicon host is stable. One way to infer the stability is to examine the spin-lattice relaxation time T_1 , which characterizes the time it takes for a spin system with a net alignment (a net magnetization) to return to its thermally equilibrated magnetization. At temperatures of about 1 kelvin, the nuclear-spin relaxation time $T_{1,n}$ (where the subscript “n” is for the nucleus) for phosphorus in silicon is expected to be longer than 10 hours (Feher 1959).

The nuclear spin qubits, however, interact with the environment through their donor electrons; as a result, the electron spin stability is also important, particularly for qubit readout (see later discussion). The electron-spin relaxation time $T_{1,e}$ (where the subscript “e” is for the electron) is approximately 1 hour at about 1 kelvin (Honig and Stupp 1960). The phase decoherence time of the electron spin, as measured by the spin-spin relaxation time $T_{2,e}$, is shorter still. Although never measured for an isolated electron system such as our qubit scheme, the $T_{2,e}$ for an ensemble of electrons was measured to be approximately 0.5 millisecond (Gordon and Bowers 1958). A recent theoretical study, appropriate for a single phosphorus donor atom in sili-

con, indicates a $T_{2,e}$ of the order of 1 second (Mozyrsky et al. 2002). This value for $T_{2,e}$ should be long enough for us to perform a quantum computation and read out the result.

The Spin Hamiltonian and Single-Qubit Operations. To understand the physics underlying the operation of the silicon-based computer, recall that phosphorus has one more electron in its outer electron shell than silicon. When it is placed into a silicon crystal lattice, phosphorus fulfills its role as a surrogate silicon atom and still has one electron left over. At very low temperatures, that “donor” electron remains bound—albeit rather loosely—to the phosphorus nucleus. The electron “talks” to the nucleus primarily through the charge (Coulomb) interaction and to a lesser degree through the hyperfine interaction, which is between the electron spin and the nuclear spin.

We exploit the hyperfine interaction to individually address single qubits. The effective low-energy, low-temperature Hamiltonian describing the spin interactions for the electron spin and the nuclear spin of an atom in the presence of a static magnetic field \mathbf{B}_0 is given by

$$H = \mu_B \mathbf{B}_0 \sigma_z^e - g_n \mu_n \mathbf{B}_0 \sigma_z^n + A \sigma^e \cdot \sigma^n, \quad (1)$$

where σ_z^e and σ_z^n are Pauli spin matrices, μ_B and μ_n are the Bohr and nuclear magnetons, respectively, and g_n is the nuclear g -factor. The hyperfine interaction is described by the term $A \sigma^e \cdot \sigma^n$.

For large values of \mathbf{B}_0 , the Hamiltonian in Equation (1) leads to a set of energy levels that correspond to the four hyperfine levels, $|0\downarrow\rangle$, $|1\downarrow\rangle$, $|0\uparrow\rangle$, and $|1\uparrow\rangle$. At the sub-Kelvin operating temperature of the computer, however, the electron spins are completely spin-polarized in the lower-energy state $|\downarrow\rangle$.

Thus, for one-qubit operations, we may ignore the electron spin polarization to a good approximation and consider only the two nuclear states $|0\rangle$ and $|1\rangle$ (Goan and Milburn 2000). The energy difference between those states is

$$\Delta E_0 = \hbar\omega_0 \cong 2g_n \mu_n \mathbf{B}_0 + 2A + 2A^2/\mu_B \mathbf{B}_0, \quad (2)$$

where ω_0 is called the nuclear resonance frequency. The resonance frequency, which is typically in the radio-frequency (rf) range, is equal to the Larmor frequency, or the rate at which the nuclear spins precess about the magnetic-field lines.

Suppose \mathbf{B}_0 is aligned along the z -axis, and the nuclear spin is initially in the $|0\rangle$ state. If we subject the spins to a secondary magnetic field that is oscillating in the x -direction at the nuclear resonance frequency, that is, a field $\mathbf{B}_1 = B_1 \cos(\omega_0 t) \hat{x}$, then the nuclear spins will begin to rotate toward the $(-z)$ -axis, or from the parallel to the antiparallel alignment (see the box “Spin Manipulation with Magnetic Resonance” on page 288). The spin rotation is equivalent to rotating a qubit in Hilbert space from the $|0\rangle$ state to some superposition of the $|0\rangle$ and $|1\rangle$ states.

As described, the \mathbf{B}_1 field will affect all spins simultaneously. To address a particular spin, we use the A -gate directly above it and modify that donor atom’s hyperfine coupling. The parameter A in Equation (1) is proportional to the magnitude of the electron probability density at the site of the nucleus, $\Psi_e(0)$:

$$A = 8/3\pi \mu_B g_n \mu_n |\Psi_e(0)|^2. \quad (3)$$

As seen in Figure 2, placing a positive voltage on the A -gate above the phosphorous atom attracts the atom’s electron cloud toward the surface and away from the nucleus, thereby reducing the magnitude of $\Psi_e(0)$. The hyperfine

energy levels of that one atom change slightly, and the resonance frequency needed to rotate the nuclear spin is reduced from, say, ω_0 to ω_- . If the frequency of the \mathbf{B}_1 field is set to ω_- , that is, $\mathbf{B}_1 = B_1 \cos(\omega_- t) \hat{x}$, then only the spin directly under the A-gate will be in resonance and will begin to rotate. Removing the voltage on the A-gate halts the rotation.

A one-qubit gate operation is therefore implemented if the silicon wafer is subjected to a transverse oscillating B-field of frequency ω_- and if the A-gate above a qubit is pulsed for a defined period. Throughout the duration of the pulse, the qubit is in resonance with the secondary magnetic field and rotates through some angle in Hilbert space. When the voltage is removed at the end of the pulse, the qubit is left in the desired superposition of the $|0\rangle$ and $|1\rangle$ states.

Two-Qubit Operations. To select adjacent pairs of qubits for two-qubit operations, we apply a positive voltage to the J-gate between them. As seen in Figure 3(a), this procedure causes the electron wave functions of the two donor atoms to overlap, and the electron spins couple together through the electron-spin exchange interaction. Because each electron is coupled to its nucleus through the hyperfine interaction, turning on the electron-spin exchange interaction also couples the nuclear spins together.

The coupled nuclear-electron spin system is fairly complex, but we can gain insight into it by looking at the effective Hamiltonian for the system:

$$H_{\text{coupled}} = H^1 + H^2 + J\sigma^1 e \cdot \sigma^2 e \quad (4)$$

This Hamiltonian is valid at energy scales that are small compared with the electron-binding energies of the donor atoms. The first two terms correspond to the hyperfine Hamiltonian—Equation (1)—of each donor, respectively, and the last term accounts for the spin exchange

Continued on page 290

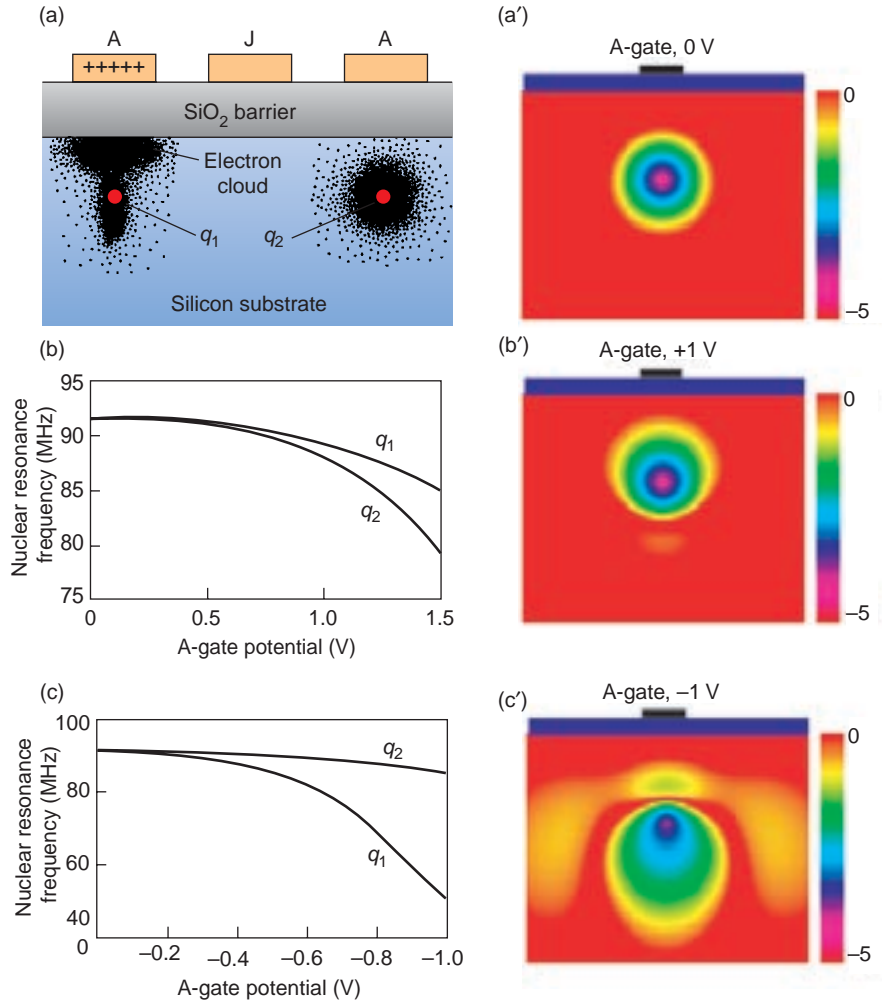
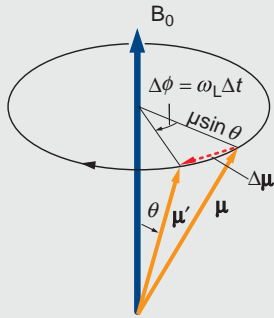


Figure 2. A-Gate Control of One Qubit

We use magnetic resonance techniques to rotate nuclear spins and place them in arbitrary superpositions of $|0\rangle$ and $|1\rangle$ qubit states. (a) In this cartoon, a small voltage is applied to the A-gate directly above a qubit. The donor electron moves away from the ^{31}P nucleus. (a') This plot of the electron probability density surrounding a donor atom with no voltage on the A-gate was obtained by solving the Schrödinger equation nonperturbatively in an isotropic effective-mass hydrogenic basis. The plot is a cross section through the nucleus, with the color variations on a logarithmic scale. The atom is buried 20 nm below the Si/SiO₂ interface. (b) The graph shows the variation of the nuclear transition frequency as a function of A-gate voltage. (b') The color plot shows the electron probability density. A positive voltage on the A-gate pulls the electron away from the nucleus, thus reducing the hyperfine coupling—the parameter A in Equation (1) in the text. In a B-field of about 2 T, the resonance frequency of a phosphorous nucleus q_1 is $\nu_0 = \omega_0/2\pi \approx 93$ MHz. With a gate voltage of 1 V, the resonance frequency of q_1 reduces to about $\nu_- = \omega_-/2\pi \approx 90$ MHz, while a neighboring nucleus q_2 is in resonance at about 87 MHz. (The proximity of the oxide barrier has a fairly large effect on the qubits, and the positive gate voltage affects q_2 more than q_1 .) Subjecting the silicon wafer to a transverse, oscillatory magnetic field of frequency ν_- would cause only q_1 to respond. (c)–(c') Initial calculations indicate that the electron probability density is more responsive to a negative gate bias, which results in better frequency discrimination between adjacent qubits.

Spin Manipulation with Magnetic Resonance



$$|\tau| = |\mu \times B_0| = \mu B_0 \sin \theta$$

$$\tau \equiv \Delta J / \Delta t = \gamma^{-1} \Delta \mu / \Delta t$$

$$|\tau| = \gamma^{-1} (\mu \sin \theta \omega_L \Delta t) / \Delta t = \gamma^{-1} \mu \sin \theta \omega_L$$

$$\gamma^{-1} \mu \sin \theta \omega_L = \mu B_0 \sin \theta$$

$$\Rightarrow \omega_L = \gamma B_0$$

Figure A. Larmor Precession
Magnetic moments precess around magnetic field lines at the Larmor precession frequency ω_L , which is derived above.

Magnetic resonance is the traditional technique for detecting and manipulating any particle, such as an electron, atom, or nucleus, that has a magnetic moment μ . The manipulation is controlled by a combination of static and oscillating magnetic fields. Classically, a particle’s magnetic moment is proportional to its angular momentum J through the relation

$$\mu = q/2m J \quad (1)$$

where q is the charge of the particle and m is its mass. Remarkably, a similar relation holds true in quantum mechanics, although we must also take into account the particle’s intrinsic spin angular momentum. In general, we can write

$$\mu = \gamma J \quad (2)$$

where the parameter γ is known as the gyromagnetic ratio. It is related to the constants in equation (1) by a dimensionless constant known as the g-factor,

$$\gamma = g (q/2m) \quad (3)$$

The magnitude and sign of the g-factor depend on the specific atom or nucleus, but are always approximately 1.

In the classical picture of a randomly oriented moment in a magnetic field $B_0 = B_0 \hat{z}$, the moment would like to lower its energy by aligning itself parallel to the applied field. But the magnetic field can only produce a torque on the moment, $\tau = \mu \times B_0$. Because the torque is directed perpendicular to the plane defined by the field, the moment does not align with the field, but like a spinning gyroscope that resists the force of gravity, precesses around the magnetic field line. By using the fact that the torque is equal to the rate of change of the angular momentum, we can derive the angular precession frequency of the moment (see Figure A):

$$\omega_L = \gamma B_0 \quad (4)$$

where ω_L is called the Larmor frequency and is measured in radians per second. Equation (4) is the single most important equation of magnetic resonance. It says that the frequency of precession about a magnetic-field line is proportional to both the magnitude of the magnetic field and the gyromagnetic ratio. Interestingly, as derived in the equations accompanying the figure, the frequency is independent of the angle θ that specifies the orientation of the magnetic moment. The Larmor frequency enables us to identify the particle because the gyromagnetic ratio is distinct for electrons and different nuclei. The Larmor frequency ($\omega_0/2\pi$) for an electron is about 28 gigahertz per tesla (MHz/T) and for a proton, roughly 45 MHz/T.

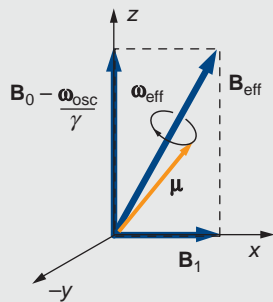


Figure B. Effective Magnetic Field in the Rotating Frame
The motion of the moment about B_{eff} is easier to describe in the frame rotating about the z-axis at the same frequency ω_{osc} as the oscillating field B_1 , since then B_1 is static. B_0 is reduced by the amount ω_{osc}/γ .

The moments precessing in the applied field can be manipulated in several ways. One common method is pulsed magnetic resonance. For a short period, we apply an oscillating magnetic field along the x-axis, $B_1 = B_1 \cos(\omega_{osc}t) \hat{x}$, where $B_1 \ll B_0$. The moment will begin to precess around a time-dependent total magnetic field consisting of B_0 plus B_1 . This complicated motion can be better understood by examining the moment in a reference frame that rotates around the z-axis with frequency ω_{osc} , the same frequency as B_1 . In the rotating frame, B_1 becomes a static field and the precession frequency about the z-axis is reduced: $\omega_L \rightarrow \omega_L - \omega_{osc}$.

Phenomenologically speaking, in the rotating frame the magnetic moment “sees” an effective field of magnitude

$$\begin{aligned} \mathbf{B}_{\text{eff}} &= \left(B_0 - \frac{\omega_{\text{osc}}}{\gamma} \right) \hat{z} + B_1 \hat{x} \\ &= \frac{1}{\gamma} (\omega_L - \omega_{\text{osc}}) \hat{z} + B_1 \hat{x} \end{aligned} \quad (5)$$

which is illustrated in Figure B. Equation (5) tells us that, when the frequency of the \mathbf{B}_1 equals the Larmor frequency, namely, at the resonance condition $\omega_L = \omega_{\text{osc}}$, the effective field has no z -component. Only the \mathbf{B}_1 field remains, and the moment will precess around the x -axis at an angular frequency ω_1 set by the magnitude of \mathbf{B}_1 , namely, $\omega_1 = \gamma B_1$. Thus in the laboratory, we can rotate a moment about the x -axis by setting the frequency of \mathbf{B}_1 to the Larmor frequency. We control the rate of rotation by adjusting the field strength and the amount of rotation by restricting the length of time that the \mathbf{B}_1 field is applied.

Pulsed magnetic resonance can be used to manipulate a qubit. Suppose a qubit state is defined by the nuclear spin orientation such that the spin aligned parallel to \mathbf{B}_0 represents the state $|0\rangle$ whereas the spin aligned antiparallel to the field represents the state $|1\rangle$. We send a current pulse through an inductive coil to create the field B_1 . If the pulse is timed to last for one-half of a precession period, or $t = \pi/\omega_1$, then the spins will rotate around the x -axis for π radians, or by 180° . If the qubit was initially in the $|0\rangle$ state, it would now be in the $|1\rangle$ state. Similarly, we can pulse the current for a time $t = \pi/(2\omega_1)$ —a so-called $\pi/2$ pulse—and rotate the qubit into an equal superposition of the $|0\rangle$ and $|1\rangle$ states, namely the state $1/\sqrt{2} (|0\rangle + |1\rangle)$. (See Figure C.)

We can also make moments rotate continuously about the x -axis. In a process known as cyclic adiabatic inversion, we sweep ω_{osc} through a range that includes the Larmor frequency. When we start the sweep, $\omega_{\text{osc}} \ll \omega_L$. According to Equations (5), there is little cancellation of the static field \mathbf{B}_0 , and \mathbf{B}_{eff} will lie substantially along the z -axis. As the frequency approaches ω_L , there is more cancellation, and \mathbf{B}_{eff} begins to rotate toward the x -axis. When $\omega_{\text{osc}} = \omega_L$, \mathbf{B}_{eff} points along the x -axis. Continuing to sweep the frequency to $\omega_{\text{osc}} \gg \omega_L$ will eventually cause \mathbf{B}_{eff} to point along the $(-z)$ -axis. If ω_{osc} is swept slowly enough (the adiabatic condition), the moments will continue to precess around \mathbf{B}_{eff} and will follow its rotation in the x - z plane from $+z$ to $-z$ (See Figure D). Reversing the sweep will cause \mathbf{B}_{eff} to rotate backwards. By continuously sweeping ω_{osc} back and forth through the resonance frequency, we effectively make the spins rotate continuously around the y -axis.

Cyclic adiabatic inversion provides one of the mechanisms by which we detect electron moments with a magnetic resonance force microscope (MRFM). A small number of moments are in resonance with \mathbf{B}_0 , \mathbf{B}_1 , and the gradient field produced by the magnetic tip at the end of the MRFM cantilever. We use cyclic adiabatic inversion to selectively rotate those moments, thus producing a tiny oscillating magnetization within the sample that in turn produces an oscillating force on the MRFM cantilever. By adjusting the rate at which we sweep ω_{osc} , we can match the forcing frequency to the cantilever’s resonant frequency, and even a small number of moments can drive the cantilever into a detectable oscillation.

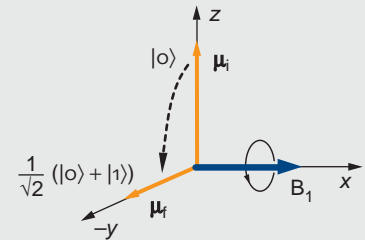


Figure C. Pulsed Magnetic Resonance

When ω_{osc} is made equal to ω_L , a moment will begin to rotate about the x -axis. We place a qubit into an equal superposition of logical states by rotating the moment through 90° with a $\pi/2$ pulse, in which B_1 is turned on for a time $t = \pi/(2\omega_1)$, where $\omega_1 = \gamma B_1$.

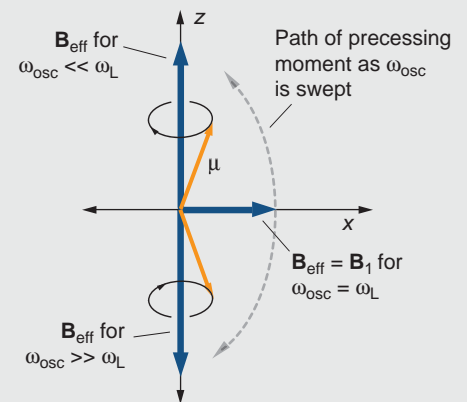


Figure D. Cyclic Adiabatic Inversion

\mathbf{B}_{eff} rotates about the y -axis when ω_{osc} is swept through the resonance frequency ω_L . If ω_{osc} changes slowly, the moment continues to precess about \mathbf{B}_{eff} and we can rotate the moment about the y -axis.

Continued from page 287

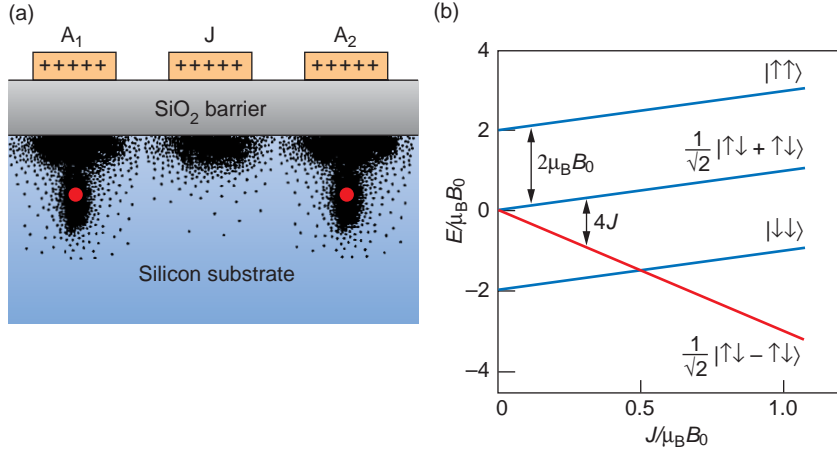


Figure 3. Coupling Two Qubits with a J-Gate

(a) When the gates A_1 and A_2 are appropriately biased, application of a small positive voltage to the J-gate lowers the potential barrier between adjacent donor sites and turns on an electron-spin exchange interaction between qubits, as seen in this cartoon. The electrons interact with the nuclei through the hyperfine interaction; hence, the two nuclear spins become coupled to each other. (b) The graph shows energy levels of the coupled electron-spin system as a function of the electron-spin exchange coefficient J , which can be modified by voltage applied to the J-gate. The electrons couple their spins to form three states with $S = 1$ (shown in blue) and one with $S = 0$ (shown in red). For $J\mu_B B_0 < 0.5$, the electrons occupy the lowest energy $S = 1$ state $|\downarrow\downarrow\rangle$. Two-qubit operations are performed in this low J regime.

interaction. The exchange coupling coefficient J is proportional to the overlap between the wave functions of the two donor electrons, so its strength is a function of the J-gate voltage.

We first examine the coupled electron-spin states by ignoring (for a moment) the contribution of the nuclear spins to H^1 and H^2 in Equation (4). The effect of the spin exchange interaction is to create coupled electron-spin states, three with total spin $S = 1$ and one with total spin $S = 0$. The respective wave functions are

$$S = 1 \quad \begin{aligned} &|\uparrow\uparrow\rangle, \\ &1/\sqrt{2} |\uparrow\downarrow + \downarrow\uparrow\rangle, \text{ and} \\ &|\downarrow\downarrow\rangle, \end{aligned}$$

$$S = 0 \quad |s\rangle = 1/\sqrt{2} |\uparrow\downarrow - \downarrow\uparrow\rangle.$$

In the absence of a magnetic field, the energy difference between the states with $S = 1$ and $S = 0$ is $4J$, an amount known as the exchange energy. In the

presence of the magnetic field B_0 that permeates the quantum computer, the $|\uparrow\uparrow\rangle$ and $|\downarrow\downarrow\rangle$ states are Zeeman-split around zero by an amount $\pm 2\mu_B B_0$, and the energies of the coupled electron-spin states vary as a function of J , as seen in Figure 3(b). Notice that the lowest-energy $S = 1$ state and the $S = 0$ state cross when the exchange energy becomes equal to the Zeeman splitting, that is, when $4J = 2\mu_B B_0$. We exploit that crossing in a qubit readout scheme discussed later.

We now consider the nuclear spin states. Conceptually, for every coupled electron-spin state, there are four possible orientations of the two nuclear spins, corresponding to the uncoupled ($J = 0$) nuclear states $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$. Thus, there are sixteen nuclear spin states in all.

Formally, we must use Equation (4) to find the energies and eigenfunctions of all sixteen.³ If we focus only on those states associated with the electron ground state $|\downarrow\downarrow\rangle$ and assume

$4J < 2\mu_B B_0$, then in order of decreasing energy, the coupled nuclear-spin states are the following:

$$\begin{aligned} &|11\rangle, \\ &|\Phi_+\rangle = 1/\sqrt{2} |10 + 01\rangle, \\ &|\Phi_-\rangle = 1/\sqrt{2} |10 - 01\rangle, \text{ and} \\ &|00\rangle. \end{aligned} \quad (5)$$

The electron-spin exchange interaction shifts the energy of the $|\Phi_-\rangle$ state below that of $|\Phi_+\rangle$ by an amount

$$\begin{aligned} \Delta E_J &= \hbar\omega_J \\ &= 2A^2 \left(\frac{1}{\mu_B B_0 - 2J} - \frac{1}{\mu_B B_0} \right), \end{aligned} \quad (6)$$

where ω_J is the nuclear exchange frequency. For $B_0 = 2T$ and $4J = 0.124$ milli-electron-volt (meV), ω_J has a value of about $(2\pi)75$ kilohertz, a frequency that approximates the rate at which binary operations can be performed on the computer.

The spin exchange interaction causes the wave functions of Equation (5) to evolve and rotate between spin states. One possible result is that the nuclear spins undergo a coordinated swapping of states: $|q_1 q_2\rangle \rightarrow |q_2 q_1\rangle$ (see the box “The Swap” on the facing page). Thus, the spin exchange interaction should automatically implement the logical two-qubit **swap** gate.

Of more interest is the **cnot** gate, which along with single-qubit operations, forms a universal set of gates from which any quantum algorithm can be executed. In the Kane system, the **cnot** corresponds to the conditional rotation of a target spin by 180° , provided the control spin is in the state $|1\rangle$. In principle, it can be realized by subjecting the wafer to a transverse

³ The energy differences between the four nuclear states associated with each electron state are very small. A graph of the nuclear-electron energy levels would look identical to Figure 3(b), except that under high magnification one would see that each line consists of four closely spaced lines.

magnetic field \mathbf{B}_1 and applying voltages to the A- and J-gates (Goan and Milburn 2000).

Suppose that the two electron spins are initially uncoupled ($J = 0$) and that the hyperfine coupling constants A_1 and A_2 of the two donor atoms are equal ($A_1 = A_2$). In that case, biasing the A-gates such that $A_1 > A_2$ distinguishes the control qubit from the target. We then turn on the spin exchange interaction ($J > 0$) and slowly make A_1 equal to A_2 . The result would be that the uncoupled qubit state $|10\rangle$ evolves adiabatically into the state $|\Phi_+\rangle = 1/\sqrt{2} |10 + 01\rangle$, and $|01\rangle$ evolves into $|\Phi_-\rangle = 1/\sqrt{2} |10 - 01\rangle$. When A_1 equals A_2 , the energy splitting between the two states is given by Equation (6). The states $|11\rangle$ and $|00\rangle$ are unaffected by the procedure.

We next apply a linearly polarized oscillating field B_1 with frequency that is resonant with the $|11\rangle - |\Phi_+\rangle$ energy difference. The field is left on until the $|11\rangle$ state has rotated into the $|\Phi_+\rangle$ state and vice versa. By executing a reverse of the sequence of steps performed at the beginning of the operation, we adiabatically transform the $|\Phi_+\rangle$ and $|\Phi_-\rangle$ states back into $|10\rangle$ and $|01\rangle$, respectively. We effect the change

$$\begin{aligned} |00\rangle &\rightarrow |00\rangle, \\ |01\rangle &\rightarrow |01\rangle, \\ |10\rangle &\rightarrow |11\rangle, \text{ and} \\ |11\rangle &\rightarrow |10\rangle. \end{aligned}$$

That is, the only qubits that are flipped are those in which the control qubit is in the state $|1\rangle$. Thus, we expect to be able to perform the **cnot** operation.

Approaches to Readout

One can evaluate the result of a quantum computation only by reading the final state, $|0\rangle$ or $|1\rangle$, of a qubit. Likewise, the ability to determine the

The Swap

Before the J-gate is turned on, the two nuclear spins are uncoupled, and each is described by the following energy eigenstates: $|\Psi_1\rangle = |00\rangle$, $|\Psi_2\rangle = |01\rangle$, $|\Psi_3\rangle = |10\rangle$, and $|\Psi_4\rangle = |11\rangle$. Once the J-gate is turned on, the coupled eigenstates are $|00\rangle$, $|\Phi_-\rangle = 1/\sqrt{2} |10 - 01\rangle$, $|\Phi_+\rangle = 1/\sqrt{2} |10 + 01\rangle$, and $|11\rangle$.

Suppose the uncoupled nuclear spins were originally in the state $|\Psi_2\rangle = |01\rangle$, and then voltage was applied quickly to the J-gate. In terms of the eigenstates of the coupled system, the system finds itself in the state

$$|\Psi_2\rangle = 1/\sqrt{2} (|\Phi_+\rangle - |\Phi_-\rangle). \quad (1)$$

The time evolution of this wave function (up to an overall phase) is given by

$$|\Psi_2(t)\rangle = 1/\sqrt{2} (|\Phi_+\rangle - e^{-i\omega_J t} |\Phi_-\rangle), \quad (2)$$

where ω_J is the nuclear exchange frequency. After a time $t = \pi/\omega_J$, the wave function will evolve into

$$|\Psi_2(\pi/\omega_J)\rangle = 1/\sqrt{2} (|\Phi_+\rangle + |\Phi_-\rangle) = |\Psi_3\rangle. \quad (3)$$

That is, the system will have evolved from the state $|01\rangle$ to the state $|10\rangle$. The spins will have swapped. If we quickly remove the voltage from the J-gate, the two-spin system will remain in the state $|10\rangle$.

state of a given qubit is critical to initializing the quantum register. Ideally, we would read the qubit state directly by measuring the donor atom's nuclear-spin state. But direct detection of a single nuclear spin is currently impossible and may forever be out of our grasp. (The strength of the coupling between a magnetic field and the nuclear spin is set by the magnitude of the nuclear magneton μ_n , which is very small.) We are therefore forced to find some other means of reading out the qubit state.

The potential solution is to correlate the nuclear spin states of a target atom with the electron spin and to find some way of determining the electron spin state. We are currently pursuing two distinct detection schemes, one involving a single electron transistor (SET) and the other, a magnetic resonance force microscope (MRFM). Both approaches require that we push the respective technologies beyond the current state of the art.

Single-Charge Measurement Using SETs. The idea behind this technique, first described by Kane (1998), is to couple the target qubit q_t to a readout qubit q_r by a J-gate, and then infer the state of q_t by monitoring the donor electrons of the coupled system. If q_t is in the state $|0\rangle$, we can cause both electrons to become localized around the readout atom (they would occupy the so-called D^- state). If q_t is in the state $|1\rangle$, each donor electron would remain bound to its respective atom. An SET would be used as an ultrasensitive electrometer to determine whether one or two electrons were bound to the readout atom.

The procedure can be understood with reference to Figure 4(a), which shows the coupled nuclear-spin states in the vicinity of the electron spin crossing. As discussed in the previous section, for $J/\mu_B B_0 < 0.5$, the lowest-energy electron spin state is the $S = 1$ state $|T\rangle = |\downarrow\downarrow\rangle$, but for $J/\mu_B B_0 > 0.5$,

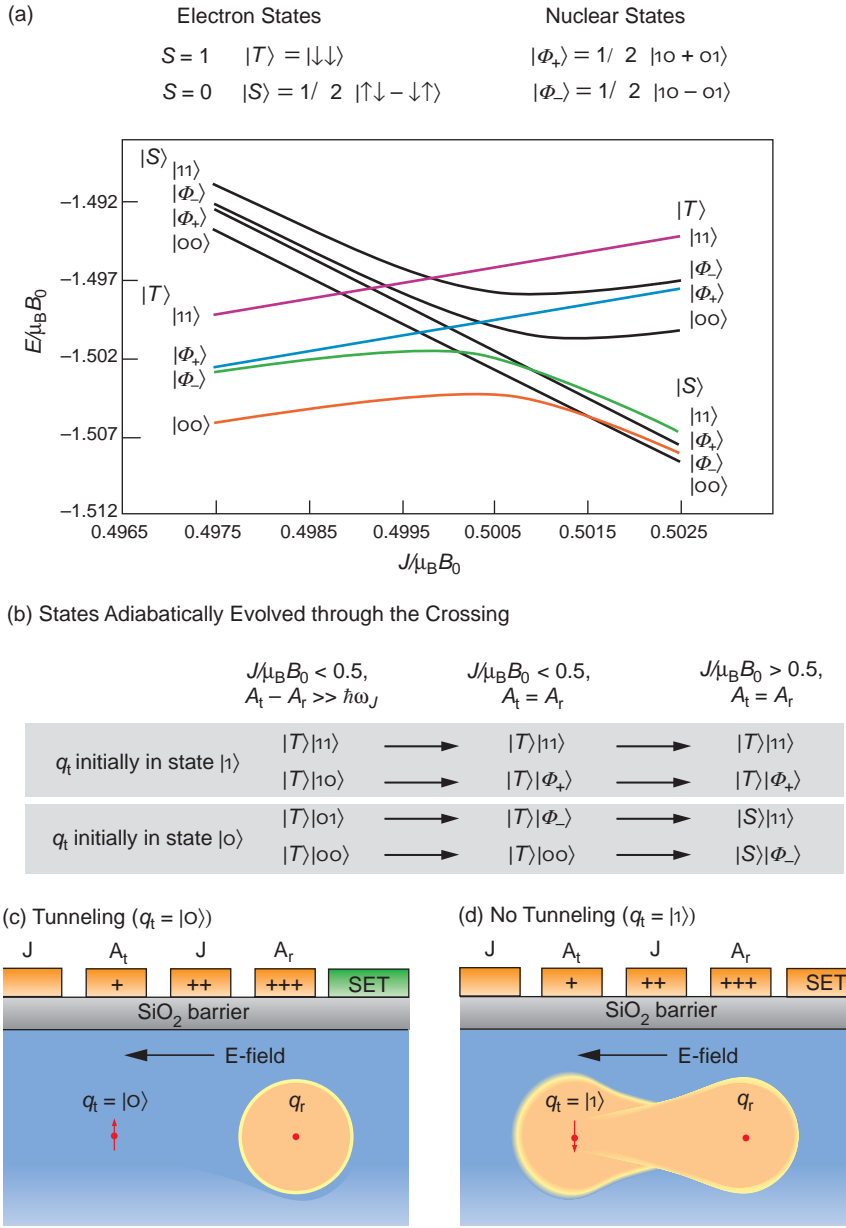


Figure 4. Single-Electron Transistor (SET) Readout Scheme

(a) The graph shows the eight lowest-energy nuclear-spin states for the coupled target and readout qubits $|q_t, q_r\rangle$ in the region where the $S = 0$ and the lowest energy $S = 1$ electron-spin states cross. (b) We can adiabatically evolve the nuclear-electron states by biasing the J- and A-gates, as seen in this (partial) sequence of steps. The electrons are initially in the $S = 1$ state $|T\rangle$. If q_t was initially in the $|1\rangle$ state, then the electrons will remain in $|T\rangle$ regardless of the state of q_r . If initially $q_t = |0\rangle$, then at the end of the sequence, the electrons will be in the $S = 0$ state $|S\rangle$. (c) Only the two electrons in the $|S\rangle$ state can bind to a single phosphorous atom in silicon. Given a suitable biasing of the gate electrodes, we can try to induce an electron to tunnel to a readout qubit q_r . If the tunneling is successful, the electrons were in the $|S\rangle$ state, and $q_t = |0\rangle$. The tunneling current would be detected by an SET located near q_r . (d) If no tunneling occurs, the two electrons were in the $|T\rangle$ state, and hence $q_t = |1\rangle$.

the $S = 0$ state $|S\rangle = 1/\sqrt{2} (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$ assumes the lower energy.

Figure 4(a) shows what happens to the eight lowest-energy nuclear-spin states as the electron-spin states cross. Focusing on the four states initially associated with $|T\rangle$, we see that after the crossing, the two higher-energy nuclear states $|11\rangle$ and $|\Phi_+\rangle$ remain coupled to $|T\rangle$, while the two lower-energy states $|00\rangle$ and $|\Phi_-\rangle$ get coupled to $|S\rangle$. In other words, as we increase J , we can adiabatically evolve both the nuclear- and electron-spin systems. If the target qubit was originally in the state $|0\rangle$, then regardless of the state of the readout qubit, the electrons will evolve into the $S = 0$ spin state. If q_t is originally in the state $|1\rangle$, the electrons will remain in the lowest energy $S = 1$ spin state. The sequence of steps, similar to those used to implement the **cnot** gate, is outlined in Figure 4(b).

We next use the fact that the only two-electron bound state of a phosphorous atom in silicon is the D^- state with total spin $S = 0$. As seen in Figure 4(c), we would bias the A- and J-gates to create an electric field between the two donor atoms. If the electrons are in the $S = 0$ state, the target electron can transfer to the readout atom, and we would know that the target atom was initially in the state $|0\rangle$.

An SET would be used to detect the presence of the second donor electron about the readout atom. In many ways, an SET is like an ordinary transistor, in that a gate electrode moderates the current flowing between a source and drain electrodes. The difference is that between the SET's source and drain lies an extremely small metallic island, which is isolated from each electrode by small patches of insulating material. The insulator acts as a tunnel junction. For current to flow, electrons must tunnel from the source to the island and then from the island to the drain. The tunneling current is greatly affected by the

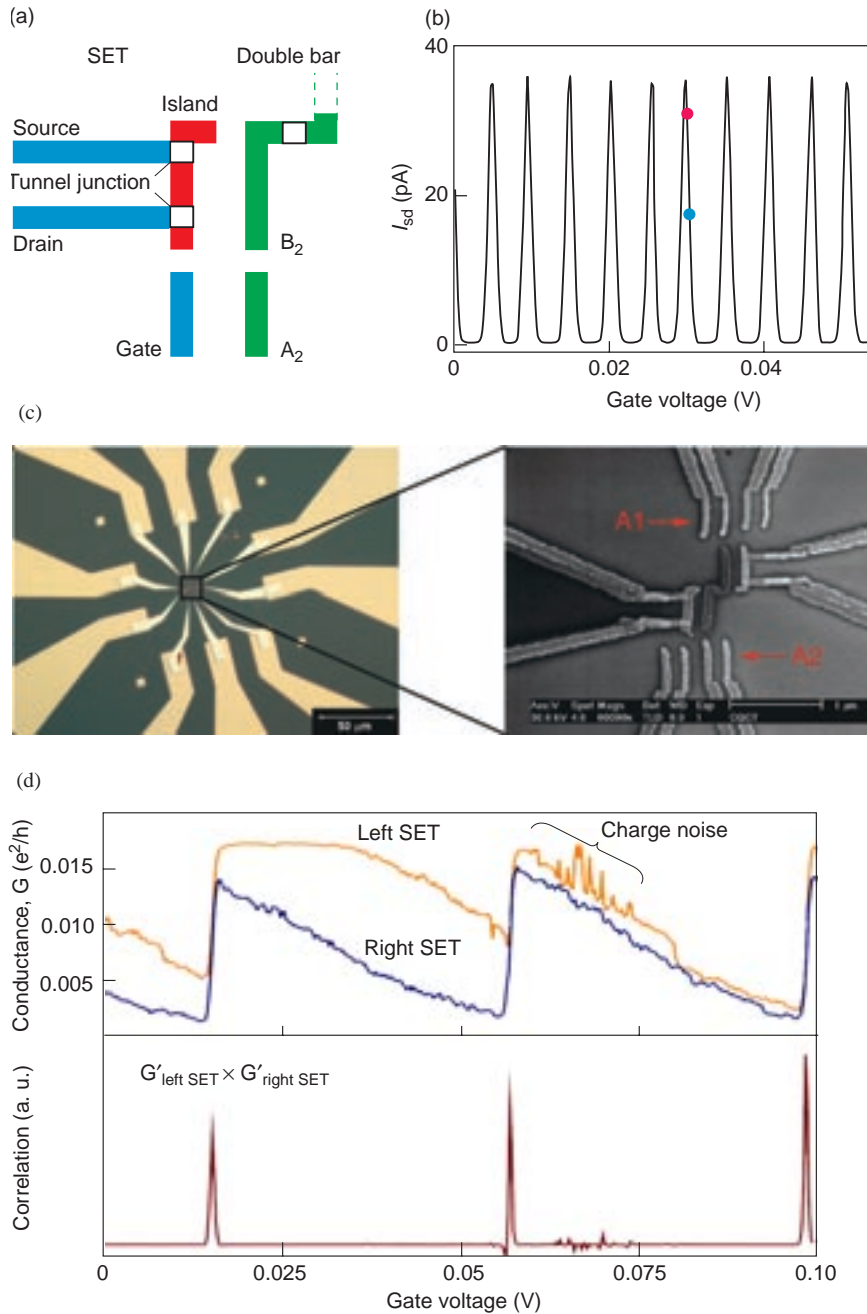


Figure 5. Twin SET Device for Readout Simulation

(a) The figure shows one half of a twin SET/ double-bar test device. The SET consists of a small metal island connected to source and drain leads by tunnel junctions and a gate electrode that is capacitively coupled to the island. Electrons can pass from source to drain only by tunneling through both junctions. The SET is located next to a metal bar B₂, which is isolated from the other bar B₁ by a tunnel junction. (b) The tunneling current I_{sd} in the SET is strongly influenced by a change in the local charge distribution. If the gate voltage is originally biased at V_0 , (blue dot), then a change in the local charge distribution effectively modifies it to $V_0 - \delta$, and the source-drain current will change dramatically (red dot). (c) This is an image of the twin-SET test device obtained with a scanning electron microscope. The image to the right is a magnified version of the central region. The twin-SET device is fabricated by a double-angle evaporation process, which replicates each of the features. Unequal voltage on A₁ and A₂ causes an electron to tunnel from one bar to the next. (d) The movement of charge is detected as a change in the source/drain conductance in both SETs simultaneously. The two signals can be correlated to discriminate the charge transfer signal from reproducible charge noise or from random noise events.

capacitive coupling between the gate and the island. This means that for particular voltage biases on the gate, source, and drain, current flow through the SET becomes exquisitely sensitive to minute changes in the charge distribution of the local environment. The presence of a single additional electron is readily detectable as a change in the SET's source/drain conductance.

We have developed several readout simulation devices to test the properties of our SETs built in house. In the device seen in Figure 5, two thin metal bars, isolated from each other by a tunnel junction, substitute for the phosphorous atoms. Control gates are used to electrostatically "push" single electrons from one bar to the next. The two SETs are then used to detect the change in the charge distribution

due to the discrete, single-electron tunneling events. Those events cause the output of both SETs to change abruptly. In contrast, signals due to unwanted charge noise (reproducible fluctuations in the conductance versus voltage curve) tend not to affect both SETs simultaneously. By correlating the outputs of the two SETs, we are able to clearly identify the single-charge transfer events and reject

spurious signals that would interfere with the readout.

Other factors, however, also need to be considered before we use an SET in a qubit readout scheme. Suppose the target qubit q_t is initially in the $|1\rangle$ state. Then, for high values of J , the coupled electrons will remain in the higher-energy state $|T\rangle$ (refer to Figure 4). This means that the coupled atomic system could lower its energy if one of the polarized electrons “relaxed” and flipped to form the state $|S\rangle$. The electron would then transfer to the readout qubit, and we would erroneously deduce that q_t was initially in the $|0\rangle$ state! Recent results suggest that the spin relaxation time is of the order of milliseconds. We must therefore pull information out of the SET on an even shorter time scale. We must be able to determine that a change occurred in the SET conductance at a time t_0 , rather than a few milliseconds after t_0 .

Unfortunately, that is difficult to do with an ordinary SET. The measurements are made at liquid helium temperatures, and the SET, sitting in a cryostat, must somehow be connected to the outside world. The capacitance of the connecting cables is fairly large, and when combined with the intrinsic resistance of the SET, produces a resistance-capacitance (or R-C) time constant for the device that is longer than the spin relaxation time. Information about the SET conductance takes too long to propagate to the outside world.

The solution to this problem is to develop a fast readout SET (Schoelkopf et al. 1998). Known as an rf SET, it is basically an ordinary SET coupled to an impedance-matching circuit that negates the effects of the external capacitance. We have recently developed a very sensitive reflection-mode rf SET that operates at 430 megahertz. It can detect the movement of a single electron in the device shown in Figure 5 in about

1 microsecond. For a system containing discrete phosphorous atoms, the readout time would likely increase to about 100 microseconds, but that is still sufficient for the readout approach discussed in this section.

The MRFM. The second approach to readout is to perform a direct measurement of the spin state of the electron surrounding the qubit and thereby infer the qubit state. To do so, we are developing an MRFM, which combines the versatility and chemical specificity of magnetic resonance with the high-resolution and ultrahigh sensitivity of an atomic force microscope (AFM). The key feature of the MRFM is that only spins contained within a defined area in the sample—the so-called sensitive slice—contribute to the detected signal. Because the location and size of that slice can be controlled, there is selective sensitivity to deeply buried structures.

Our MRFM, developed at Los Alamos in collaboration with Michael Roukes of Caltech, is illustrated in Figure 6. The microscopic, sharp-pointed magnetic tip is bonded to the end of a tiny cantilever. As in an ordinary AFM, the tip is scanned over a sample, and signals are recorded at every point. In our instrument, however, the magnetic field from the tip $\mathbf{B}(\mathbf{r})$ interacts with all the electron spins in the substrate through the magnetic gradient force, $\mathbf{F}(\mathbf{r}) = (\mathbf{m} \cdot \nabla)\mathbf{B}(\mathbf{r})$, where \mathbf{m} is the net magnetization of the spins. Depending on the spin orientation, the force on the tip is either repulsive or attractive. The net orientation of the electron spins in the sample, therefore, causes a tiny deflection of the cantilever.

We interact with only a subset of the spins through magnetic resonance. The sample is immersed in a static magnetic field $\mathbf{B}_0 = B_0 \hat{z}$, so the precession frequency of the spins around the magnetic-field lines is proportional to B_0 plus the z -component of $\mathbf{B}(\mathbf{r})$,

that is, the total magnetic field in the z -direction. Magnetic resonance comes into play when we subject the spins to an oscillating magnetic field \mathbf{B}_1 that is aligned in the x -direction. Because the magnitude of $\mathbf{B}(\mathbf{r})$ decreases rapidly with distance, only those spins that are at the correct distance from the tip are in resonance with the \mathbf{B}_1 field. Spins that are too close to the tip will have a higher resonant frequency; those that are farther away, a lower frequency. Thus, for a given field gradient and a fixed cantilever position, a resonance frequency becomes correlated with positions inside the sample. With reference to Figure 6, all spins that lie within a small, hemispherical shell beneath the tip (the sensitive slice) have the same resonance frequency.

To detect those spins, we use the technique of cyclic adiabatic inversion, discussed in the box “Spin Manipulation with Magnetic Resonance” on page 288. In essence, we continuously rotate the selected spins at the resonant frequency of the cantilever. The continuous up and down reorientation of the spins creates an oscillating force on the tip that amplifies the cantilever’s natural up-down motion. The situation is analogous to pushing a child’s swing at its natural frequency of oscillation: with each push, the amplitude of the motion becomes larger. After thousands of spin rotations, the amplitude of the cantilever’s up-down motion has increased by about an angstrom, which is large enough to be detected with a fiberoptic interferometer. The fiber sits slightly above the back of the cantilever, and laser light sent down the fiber interferes with itself as it reflects from both the cantilever and the fiber’s end. By monitoring changes in the interference pattern, we can detect the oscillations.

The orientation of the nuclear spins can be inferred from the frequency at which the electron spin resonance

occurs. Because of the hyperfine interaction, the resonance frequency of an electron spin flip depends on the nuclear-spin state. Considering the hyperfine states of a single qubit, the $|\downarrow\downarrow\rangle$ to $|\uparrow\downarrow\rangle$ transition has a different energy than the $|\downarrow\uparrow\rangle$ to $|\uparrow\uparrow\rangle$ transition, and thus there are two resonance frequencies for an electron spin transition. Measurement of, say, the higher resonance frequency would correspond to the nuclei in the sample being aligned with the \mathbf{B}_0 field.

The discussion so far has centered on detecting many nuclear spins, but to read out the result of a quantum computation, we need to measure a single nuclear spin. That such measurement is at all possible is due to the exceedingly high spatial resolution of the MRFM, which is determined by the thickness Δz of the hemispherical shell. The thickness is inversely proportional to the magnitude of the field gradient:

$$\Delta z \cong \frac{\Delta\omega_r}{\gamma|\nabla_z \mathbf{B}(\mathbf{r})|}, \quad (7)$$

where γ is the gyromagnetic ratio and $\Delta\omega_r$ is the linewidth of the resonant electron-spin transition that is being driven by the MRFM. For phosphorus atoms in silicon, $\Delta\omega_r/\gamma$ is on the order of 1 milligauss. A field gradient of about 10^5 tesla per meter (T/m) will then produce a thickness that is much less than 1 angstrom, even when the hemispherical shell extends several hundred angstroms beneath the substrate surface. In that case, the sensitive slice would be so thin that only a single donor electron would be in resonance with the MRFM probe.

We have conducted numerous experiments to measure the field gradient of our specialized magnetic tips. With the tip about 2 micrometers from the surface, we have measured a field gradient approaching 10^4 T/m (see Figure 7). From this value, we estimated Δz and the volume of our hemi-

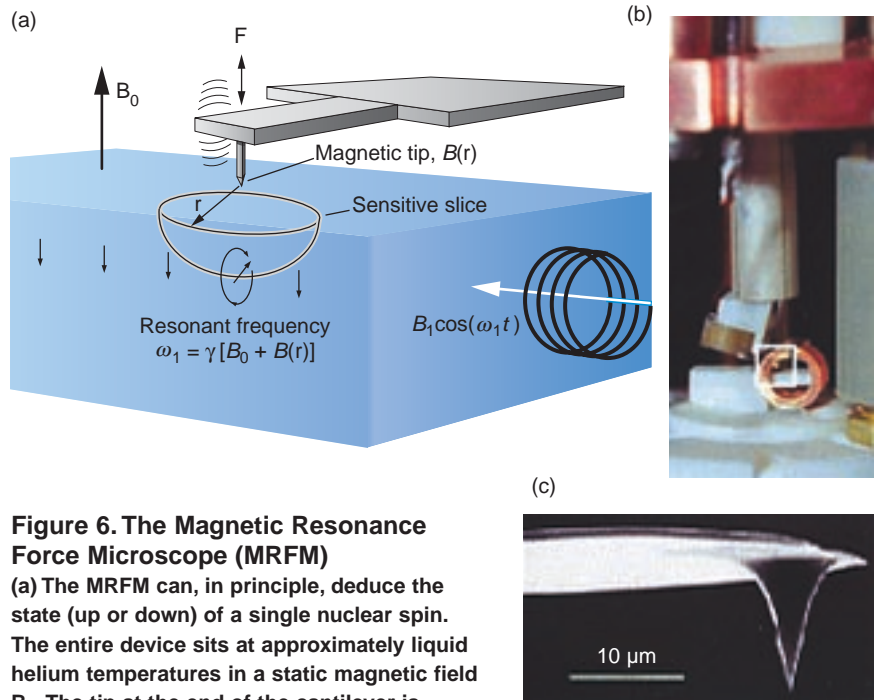


Figure 6. The Magnetic Resonance Force Microscope (MRFM)

(a) The MRFM can, in principle, deduce the state (up or down) of a single nuclear spin. The entire device sits at approximately liquid helium temperatures in a static magnetic field \mathbf{B}_0 . The tip at the end of the cantilever is coated with a magnetic material that generates a magnetic field $\mathbf{B}(r)$ that changes rapidly with the distance r . The interaction between the electron spins in the sample and the magnetic field gradient due to $\mathbf{B}(r)$ produces a force that deflects the cantilever. We interact with only a small subset of spins, located within a hemispherical shell of radius r_1 , by subjecting the sample to an oscillating magnetic field $B_1 \cos(\omega_1 t)$, where $\omega_1 = \gamma[B_0 + B(r_1)]$. By using the technique of cyclic adiabatic inversion, we can cause the spins to oscillate between the up and down states at the cantilever resonance frequency, thus driving the cantilever into measurable oscillation. We detect the oscillation with an optical device. The electron-resonance frequency can then be correlated with a nuclear spin orientation. (b) The MRFM tip assembly and sample mount are shown in this photo. The vertical tube is a piezo scanning tube, which moves the tip over the sample, while the circular feature is the induction coil that produces B_1 . The white box highlights the magnetically coated tip, shown under high magnification in (c).

spherical shell. Then, knowing the spin density of the sample, we estimated the number of spins that contribute to the signal. For the data shown in Figure 7, the number is between one thousand and ten thousand electron spins.

Because the field gradient increases nearer to the tip, sensitivity should be greater if the tip is closer to the surface. But mechanical and thermal noise also deflect the tip and cantilever. As we begin to interact with fewer spins, the “signal” force due to spins eventually becomes less than the “noise” force due to

unwanted sources. By equating expressions for the signal force to the noise force, we can derive an expression for the minimum detectable magnetic moment, m_{\min} , needed to give a signal to noise of 1:

$$m_{\min} = \frac{1}{|\nabla_z \mathbf{B}(\mathbf{r})|} \sqrt{\frac{2kk_B T \Delta\nu}{\pi Q f_c}}. \quad (8)$$

In Equation (8), k_B is the Boltzmann constant; T , the temperature; and $\Delta\nu$, the detection bandwidth. The other parameters describe the cantilever: its force constant k , resonant frequency f , and quality factor Q . The three key

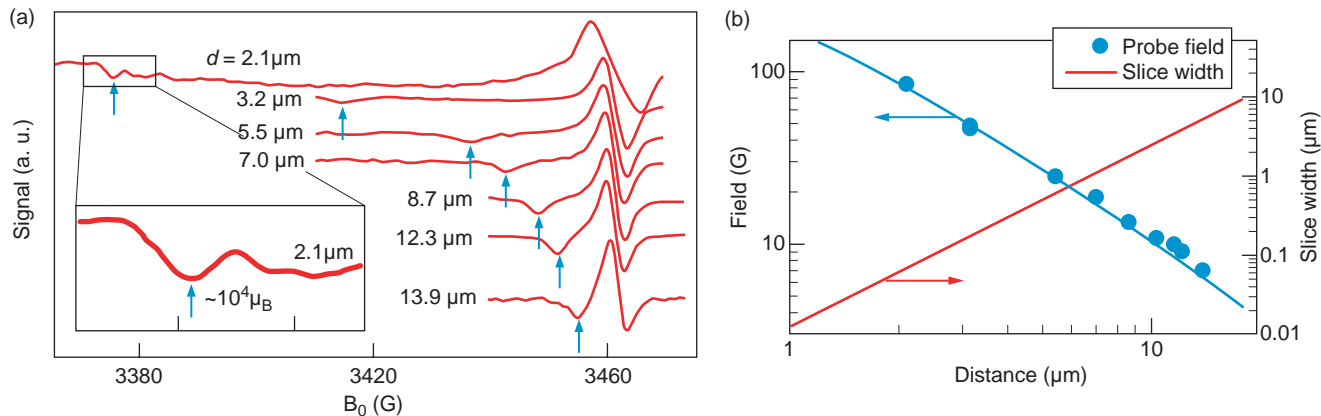


Figure 7. Sensitivity of the MRFM

(a) If the magnetic tip is kept at a fixed distance d from the sample, then lowering the value of the magnetic field corresponds to sweeping the sensitive slice upwards, toward the surface. At some point, the slice leaves the sample, and the resonance condition changes dramatically. That change is seen in the derivative of the MRFM signal as a dip, indicated

by the arrows. (b) By following the dip as a function of tip height, we can measure the tip's magnetic field. From the field gradient, we then calculate the width of the sensitive slice using Equation (7) in the text. Knowing the spin density within the sample, we use the slice width to deduce how many spins produced the signal and thereby infer the MRFM sensitivity.

parameters that we can optimize are the field gradient, temperature, and quality factor.

We believe that the sensitivity of the MRFM is currently limited by surface contamination on the sample. As the tip approaches the surface, the contamination acts like a viscous force that damps the oscillatory motion—that is, it lowers the Q in Equation (8). To solve this problem, we are upgrading the equipment so that the sample be transferred from a surface preparation chamber into the microscope without leaving the ultra-high vacuum environment. The system will also be cooled to temperatures between 250 and 300 millikelvins in a helium-3 dilution refrigerator, a technique that is compatible with maintaining the sample under ultrahigh vacuum.

Detection of a single electron moment requires that

$$m_{\min} = 1 \mu_B \cong 10^{-23} \text{ joule/tesla} . \quad (9)$$

Given a field gradient of 10^5 T/m, the signal force on the cantilever is

approximately 10^{-18} newton (the weight of approximately two million phosphorus atoms). We believe that an upgraded, low-temperature microscope will allow us to observe the magnetic resonance signal of a single electron spin.

SSQC Fabrication Progress

Implementing our quantum-computing scheme requires that we produce a very regular array of phosphorus atoms in pure silicon, in which each donor is located precisely beneath a metal A-gate on the surface. The spacing between adjacent phosphorus donors is chosen to ensure that the electron-spin exchange interaction is minimal when there is no voltage on the J-gate lying between the donors. We want the two electron wave functions to overlap, but only slightly. Calculations (Goan and Milburn 2000) indicate that a separation of 10 to 20 nanometers between donors is required.

A nominal donor spacing of

20 nanometers translates into gate structures that are less than 10 nanometers in width. Fabricating a highly regular metal array on that scale, even with state-of-the-art techniques, is at the limits of the electron-beam techniques used in making conventional electronics. That problem, however, pales when compared with the difficulties we face in making a precisely aligned array of phosphorous donors that is buried under layers of silicon. The difficulties have led us to pursue two different fabrication strategies, known as the top-down and bottom-up approaches.

In the top-down approach, phosphorous atoms will be implanted by ion bombardment into specific sites on the silicon wafer. Because the ion scatters as it slows down in the silicon, we will not know the exact location of the donors, only that they will lie within close range of the defined implantation area. The top-down approach provides a rapid means to demonstrate proof of principle and allows us to fabricate a two-donor device that can be used to test readout

strategies and, possibly, quantum operations. Scaling this approach to large numbers of qubits will be challenging, because of the irregular spacing of the donor array.

In contrast, the bottom-up approach will use a scanning tunneling microscope (STM) with which to place phosphorous atoms on a clean silicon surface in a precisely arranged array. The array will then be overgrown with silicon, and the gate structures will be laid down by electron beam lithography (EBL). This approach, although more difficult to implement, could in principle allow us to build a Kane-type computer with the required precision. The bottom-up approach is not discussed here but is described in detail in the article “Fabricating a Qubit Array with a Scanning Tunneling Microscope” on page 302.

Top-Down Approach for Creating a Two-Donor Device. A host of issues surrounds the operation and readout of the nuclear-spin quantum computer. A key concern involves the transfer of the electron from the target to the readout atom during readout. The two electrons in the D^- state are not bound very strongly to the phosphorous atom, and the electron may be lost during the transfer. The initial phosphorus-phosphorus (P-P) system would transform into a $P-P^+$ system. If that is the case, we may need to use extra electrodes in order to create a deeper potential that will confine the electron or to employ a different readout atom (such as tellurium) that has a more strongly bound two-electron state.

Our current goal for the top-down approach is to produce a device that can be used to study the controlled electron transfer between two donors. We intend to ionize one of the two phosphorus atoms and then study the coherent transfer of the remaining electron between the two donor atoms

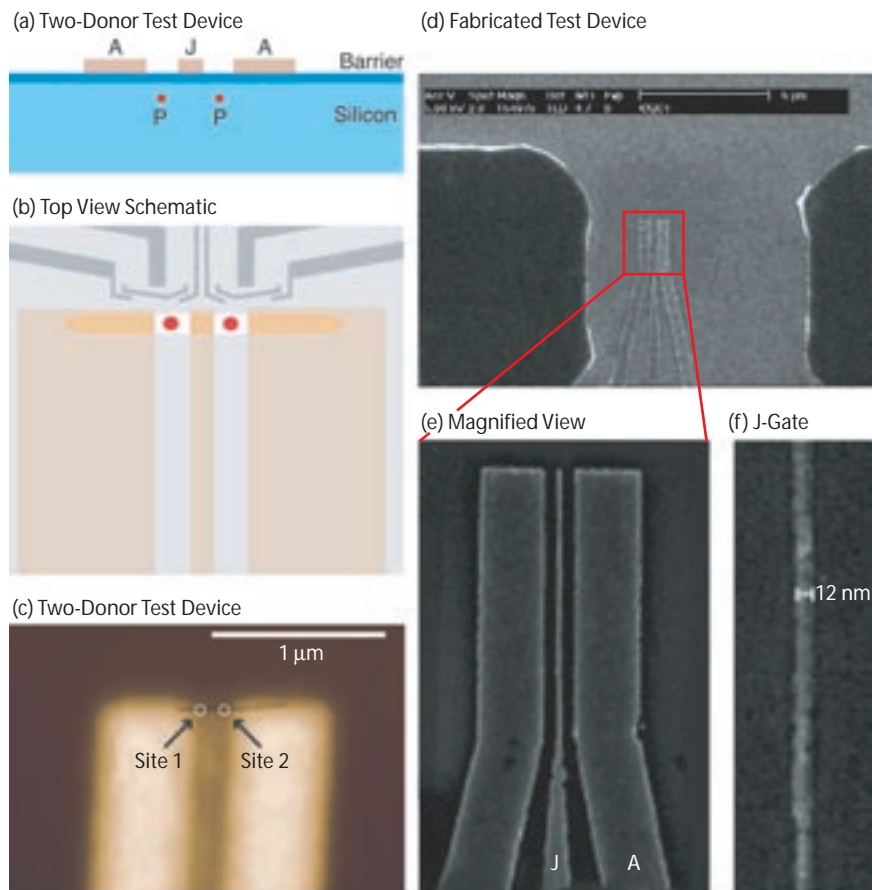


Figure 8. Fabrication of a Two-Donor Device

(a) For proof-of-principle experiments involving the transfer of an electron between two phosphorus donors, it is not necessary to configure A-gates above the donors. Instead, we are configuring the three-gate device shown in side view in this illustration. (b) The top view schematic shows that the donors will be implanted between the A- and J-gates (red circles). The single horizontal line at the top of the gates represents an opening in a polymer resist layer. Ions can enter the substrate only where the line crosses the gates. The schematic also shows representative SET devices located near the implantation sites. (c) AFM image of an actual device prior to implantation. The narrow horizontal line near the end of the gates is a 20-nm-wide opening in the resist. (d) An SEM image of a fabricated metal-gate array is shown. The large metal structures on either side of the gates are aluminum electrodes used to detect the impact of ions during implantation. (e) This magnified view shows the central A-, J-, and A-gates. We have fabricated gate arrays with J-gate widths of less than 15 nm and gate separations down to 30 nm. The image in (f) shows a J-gate made from a titanium/gold alloy that is only 12 nm wide.

in the $P-P^+$ system. For this purpose, we have relaxed the stringent constraints of the Kane computer architecture and designed the simple device shown in Figure 8. It consists of two A-gates separated by a single J-gate. Two phosphorous atoms will be

implanted between the A- and J-gates, an arrangement that is sufficient for charge transfer experiments.

Device fabrication starts with a wafer that is already topped with a barrier layer of SiO_2 . To deposit metal A- and J-gates on the surface

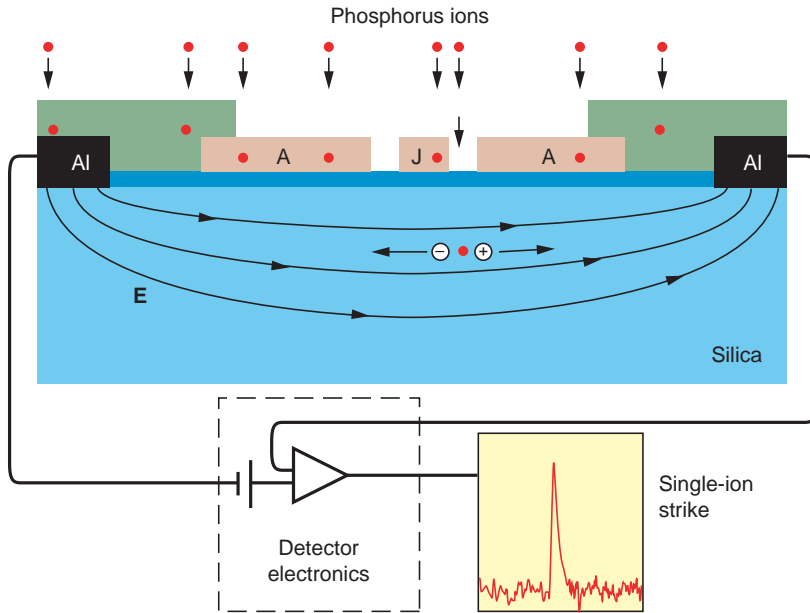


Figure 9. Detection of a Single-Ion Impact

We can detect in situ the impact of a single phosphorous ion in the silicon wafer. Aluminum electrodes are deposited on either side of the implantation site. An electric field applied between the two electrodes is used to separate electron-hole pairs that are created by ion impact. Each type of charge carrier migrates to its respective electrode and produces a current pulse that is detected by an external circuit. The implantation is halted after two such pulses have been detected.

before ion implantation, we use EBL techniques. A new layer of resist is created that covers the entire surface, including the gates. A second EBL exposure then patterns a thin line across the gates. This pattern is developed so that two tiny channels, each approximately 15×30 nanometers, are created on each side of the J-gate. The channels extend down to clean SiO_2 and define the implantation sites.

Next, we bombard the wafer with phosphorous ions. Although most of the ions are stopped in the mask, some go through the channels, strike the wafer, and get implanted about 10 nanometers below the Si/SiO_2 interface. After implantation, the device is heated to between 900°C and 950°C to anneal any damage to the silicon lattice. As a final step, we lay down the SETs. Creating the SETs

after the anneal (instead of making them in the same step as the control gates) protects their fragile tunnel barriers, which would likely be degraded should they be submitted to temperatures above 900°C .

Because we want only one phosphorus atom per implantation site, the key to this entire process is the ability to detect a single ion after it has struck the silicon. And it is the properties of the silicon itself that help us fulfill this task. The energetic phosphorous ion produces a cascade of electron-hole pairs as it slows down and comes to a stop in the silicon matrix. Those charge species can be separated by an applied electric field, accelerated, and detected as a current pulse in an external circuit (see Figure 9). Voltage applied to surface electrodes straddling the implantation sites produces the field and transmits

the pulses. The intrinsic silicon substrate makes this in situ particle-detection system highly efficient because the accelerating electric field extends fully between the two electrodes. We have demonstrated detection efficiencies of over 99 percent. Unfortunately, we cannot tell where the ion falls, and as there are two holes, there is only a 50 percent chance of creating a two-donor device with one donor in each hole. Although there are ways to improve those odds (by masking all but one hole with a specialized AFM cantilever, for example), in the short term, a 50 percent success rate is acceptable.

So far, no one has come close to seeing the transfer of a single electron between two precisely located, nanofabricated donor atoms. We hope to do so with a device similar to the one described above by September or October of 2002. We would then be in a position to study the coherent transfer of the electron between two donor atoms and possibly obtain information on decoherence mechanisms of relevance to spin readout.

Unlike the simple test device shown in Figure 8, an ideally configured device would have the A-gates directly above the phosphorous donors. We have designed a process to fabricate such a device (McKinnon et al. 2001). A multilayer electron-beam-sensitive resist is deposited on top of a SiO_2 -coated silicon wafer, the resist is partially developed, and a linear array of ion-implantation channels is patterned in the resist with EBL techniques. The wafer is bombarded with phosphorous ions. The resist is then fully developed, and triple-angle metal evaporation is used to deposit the metal gate array and the SETs on the SiO_2 surface. Because only one mask is used to define the location of both the ion implantation sites and the gates, the A-gates are automatically registered over the qubits.

Figure 10 shows a six-donor test

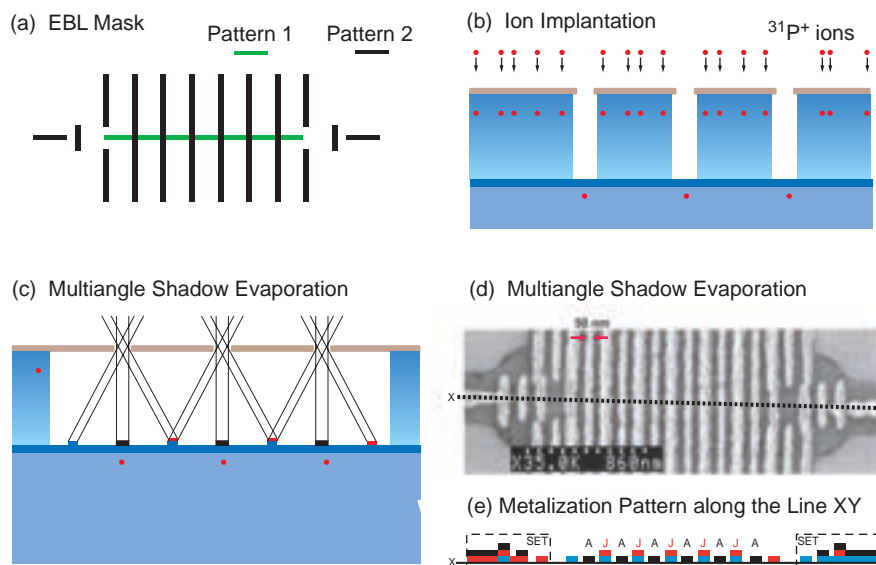


Figure 10. Creating a Multidonor Device

We have developed a process for creating gates directly over implanted ions.

(a) In separate steps, patterns 1 and 2 of this EBL mask are partially developed in a trilayer resist that coats the substrate. The resist then sustains a series of lines with tiny channels that extend down to the substrate, where patterns 1 and 2 intersect.

(b) A cross section of the resist after partial development shows the channels.

The wafer is now bombarded with phosphorus ions. Some ions make it to the silicon surface and are implanted 5–10 nm below the Si/SiO₂ interface. (c) The resist is fully developed and material is removed, leaving behind a large cavity between the SiO₂ surface and the self-supporting top layer. Triple-angle shadow evaporation is then used to lay down an array of metal gates on the SiO₂ surface. (d) This photo shows a potential six-donor device with two readout SETs. No ions were implanted into this device. (e) A schematic side view of the device reveals the metalization pattern that results from the triple-angle shadow evaporation process.

device made by triple-angle evaporation. No ions were implanted into this device, and as can be seen in the figure, the triple-angle process does not yet result in gates that are sufficiently narrow to allow implementing the Kane scheme. We also have to address the problem of maintaining the integrity of the SET through the high-temperature anneal. However, the fundamental idea is robust.

The precision of the donor arrays produced by the top-down approach will be limited by straggling, which is inherent to ion implantation, and by the diffusion of dopants during the annealing step. Recent calculations indicate that small irregularities in the ion array could impair the operation

of the quantum computer. That is why we are also pursuing the bottom-up fabrication approach, which might lead to a device with a very regular, well-characterized donor array.

Concluding Remarks

Phosphorous in silicon is a very clean, well-understood solid-state system. In its turn, NMR is a very well understood nuclear-spin manipulation technique. Performing NMR on a silicon chip implanted with phosphorus can therefore make for a very powerful quantum computer.

But the creation of a silicon-based solid-state computer presents such an

enormous technical challenge that we must explore several strategies for building and implementing almost every aspect of the device. Hence, we investigate both SETs and magnetic resonance force microscopy as a means to read out the qubit state. Similarly, we have pursued two complementary fabrication strategies: the top-down process, which uses industrial production techniques, such as ion implantation and EBL, to produce a few-qubit device, and the bottom-up process, which involves advanced STM techniques and conventional molecular-beam epitaxy. Although the bottom-up approach is less suited to high-throughput production, it has the potential of leading to large, highly regular qubit arrays. We have made significant progress along all these parallel development paths.

Currently, scaling up a solid-state computer to over a million qubits is a goal that appears so distant as to be nearly out of sight. Yet less than fifty years ago, computer companies attempting to reduce the size of their machines were just becoming aware of a new strategy known as integrated circuits. Those early chips were crude and contained but a few transistors, but from them, evolved the modern, densely packed integrated circuits of today. Like those early chips, the quantum devices developed so far are rudimentary. No doubt, the challenges we face in building a real silicon-based quantum computer are significant, but our initial results offer hope that large-scale quantum computing may one day be realized. ■

Acknowledgments

Building a silicon-based solid-state quantum computer is a difficult and challenging project that cannot be achieved without the efforts of a large research team. At the University of New South Wales, Tilo Buehler,

Rolf Brenner, David Reilly, Bob Starrett, and Dave Barber have worked in the development of SET and rf SET readout schemes. Similarly, important contributions to the top-down fabrication approach have been made by Fay Stanley, Eric Gauja, Nancy Lumpkin, Rita McKinnon, Mladen Mitic, Linda Macks, and Victor Chan. At the University of Melbourne, Cameron Wellard has supported the device modeling; Changyi Yang, Paul Spizzirri, and Robert Short have worked on single-ion implantation technologies; and Jeff McCallum and Steven Prawer have contributed in a range of areas. The Magnetic Resonance Force Microscope research would not have been possible without the contributions of Denis Pelekhov, Michael Roukes, Andreas Suter, and Zhenyong Zhang.

Further Reading

- Feher, G. 1959. Electron Spin Resonance Experiments on Donors in Silicon. *Phys. Rev.* **114** (5): 1219.
- Goan, H., and G. Millburn. 2000 (February). Silicon-Based Electron-Mediated Nuclear Spin Quantum Computer. CQCT, University of Queensland internal report.
- Gordon, J. P., and K. D., Bowers. 1958. Microwave Spin Echoes from Donor Electrons in Silicon. *Phys. Rev. Lett.* **1**: 368.
- Honig, A., and E. Stupp. 1960. Electron Spin-Lattice Relaxation in Phosphorus-Doped Silicon. *Phys. Rev.* **117** (1): 69.
- Kane, B. E. 1998. A Silicon-Based Nuclear Spin Quantum Computer. *Nature* **393**: 133.
- McKinnon, R. P., F. E. Stanley, T. M. Buehler, E. Gauja, K. Peceros, L. D. Macks, M. Mitic et al. 2002. Nanofabrication Processes for Single-Ion Implantation of Silicon Quantum Computer Devices. (To be published in *Smart Mater. Structures*.)
- Mozyrsky, D., Sh. Kogan, and G. P. Berman. 2001. Time Scales of Phonon Induced Decoherence of Semiconductor Spin Qubits. [Online]: [http://eprints.lanl.gov \(cond-mat/0112135\)](http://eprints.lanl.gov (cond-mat/0112135)).
- Schoelkopf, R. J., P. Wahlgren, A. A. Kozhevnikov, P. Delsing, and D. E. Prober, 1998. The Radio-Frequency Single-Electron Transistor (RF-SET): A Fast and Ultra-sensitive Electrometer. *Science*, **280**: 1238.

Robert G. Clark received Bachelor of Science and Ph.D. degrees in physics from the University of New South Wales in Sydney, Australia, in 1973 and 1983, respectively. Between 1984 and 1990, Bob was a lecturer in physics at the University of Oxford, England, and research group head at Clarendon Laboratory at Oxford. During this period, he received a number of awards and distinctions, including the Wolfson Award for prestigious research (1988). In 1990, Bob became professor of experimental physics at the University of New South Wales. He is now heading the Center for Quantum Computer Technology in Sydney, Australia, and is leading the Australian effort for developing a silicon-based solid-state quantum computer.



P. Chris Hammel received his B.A. from the University of California at San Diego and his Ph.D. from Cornell University (1984). Following a postdoctoral appointment at the Massachusetts Institute of Technology, Chris came to Los Alamos National Laboratory on a J. Robert Oppenheimer Fellowship. He became a staff member at Los Alamos in 1989. He is a Fellow of Los Alamos National Laboratory and was awarded the Los Alamos Fellows Prize in 1995. He is also a Fellow of the American Physical Society. In June, 2002, Chris joined the Physics Department at the Ohio State University as an Ohio Eminent Scholar. In addition to his current research, which is focused on the development of an ultrasensitive magnetic resonance force microscope, he is actively engaged in magnetic resonance studies of correlated electron systems including high-temperature cuprates and heavy fermions.



Andrew Dzurak is a program manager at the Centre for Quantum Computer Technology and Associate Professor within the School of Electrical Engineering at the University of New South Wales, Sydney. He has collaborated with researchers at Los Alamos since 1996. Andrew has 15 years of experience in the fabrication and cryogenic electrical measurement of quantum effects in semiconductor nanostructures, beginning with a Ph.D. at the University of Cambridge (the United Kingdom). Andrew moved to UNSW to take up an Australian Postdoctoral Fellowship and has worked closely with Robert Clark to establish the Semiconductor Nanofabrication Facility, of which he has been Deputy Director since 1995. Within the Centre, Andrew manages research in integrated devices for the control and readout of phosphorous qubits in silicon.



Alex Hamilton received his Ph.D. from the University of Cambridge, the United Kingdom, in 1993. He is a senior lecturer in physics in the School of Physics at the University of New South Wales, as well as manager for the Quantum Measurement Program at the Centre for Quantum Computer Technology. He is interested in the study of the fundamental properties of submicron electronic devices, with particular interest in the regime in which strong interactions between adjacent devices influence the behavior of the complete system.



Lloyd Hollenberg is a senior lecturer at the University of Melbourne, and since 2001 has been manager of the Device Modeling Program at the Centre for Quantum Computer Technology. He received his Ph.D. from the University of Melbourne in 1989. His interests include the general aspects of quantum many body systems, and in particular field theoretic and spin systems. At present, his work focuses on the physics of the Kane quantum computer and on quantum algorithms.



David N. Jamieson is an associate professor and reader at the University of Melbourne, Australia. He received his Ph.D. from the University of Melbourne, and has been Director of the Microanalytical Research Centre (MARC) of the School of Physics, University of Melbourne since 1996. In 2000, he joined the Centre for Quantum Computer Technology as the manager of the Ion Beam Program. He aims to adapt the technology of single ion detection and control to the construction of phosphorous arrays in silicon.



Christopher I. Pakes received his Ph.D. in physics in 1999 from the College of Physics, Birmingham, the United Kingdom. He has studied metrological applications of nanotechnology and spin-polarized properties of oxide functional materials, using ultrahigh vacuum, low temperature STM and dc SQUIDs. In 2000, Chris joined the Center for Quantum Computer Technology and was appointed manager of the Atomic Level Manipulation and Imaging Program in 2001.





Fabricating a Qubit Array with a Scanning Tunneling Microscope

*Marilyn E. Hawley, Geoffrey W. Brown,
Michelle Y. Simmons, and Robert G. Clark*

The Australian Centre for Quantum Computer Technology and Los Alamos National Laboratory are working together to answer the question, “Can the solid-state quantum computer (SSQC) proposed by Bruce Kane (1998) be built?” Illustrated in Figure 1, the architecture put forward by Kane requires a linear array of phosphorus atoms (nuclear spin $1/2$) inside an isotopically pure silicon-28 (spin 0) wafer. The spacing between the atoms needs to be about 20 nanometers, and the array will be located 5 to 20 nanometers beneath the silicon surface. An array of metal electrodes, isolated from the silicon by a thin insulating layer of silicon dioxide (SiO_2), will sit above the qubit array and needs to be precisely registered to it. Because the array is so small and because the silicon overlayer must be nearly free of impurities and crystalline defects for the computer to operate properly, we must achieve unprecedented control of the fabrication process.

Our efforts to build the SSQC focus on a novel “bottom-up” fabrication approach. Starting with a clean silicon surface, we will build each layer of the device in succession, first creating the phosphorus array and embedding it in the surface, then growing the silicon overlayer, the SiO_2 insulating layer, and finally laying down the metal electrodes. (We are also pursuing a “top-down” fabrication approach, which along with information about the operating principles of the computer, is described in the article “Toward a Silicon-Based Nuclear-Spin Quantum Computer” on page 284.)

The scanning tunneling microscope (STM) plays a central role in the bottom-up approach, serving as both a fabrication and electrical characteriza-

tion tool. To create the phosphorus array, we employ STM-based hydrogen lithography, developed by Joseph Lyding’s group at the University of Illinois at Urbana-Champaign (Lyding et al. 1994). Immediately following the array fabrication step, the silicon overlayer will be grown by molecular beam epitaxy to encapsulate the array. Our STMs have variable temperature control so that we can anneal the overlayer in situ, and thus be in a position to study the stability of the phosphorus array during silicon overgrowth. We can also identify potential defects and impurities that could impair computer operation. Once the thin SiO_2 layer is grown, we will create the metal-gate array using state-of-the-art electron beam lithography (EBL) technology.

In this article, we summarize our progress in building the phosphorus array, overgrowing the silicon layer, and checking whether the latter step alters the array. To convey the central role of the STM in building the SSQC, we start by presenting the principles that make it such a powerful fabrication and characterization tool.

Scanning Tunneling Microscopy

The STM probes the surface of a sample by inducing electrons to tunnel between the surface and the tip. As illustrated in Figure 2, an extremely sharp metallic tip (with radius of curvature R that is typically about 10 nanometers) is brought to within a few angstroms of a sample’s surface. The thin vacuum region separating the tip and the sample forms a potential barrier, and a bias voltage between the tip and the sample causes more electrons to tunnel through the barrier

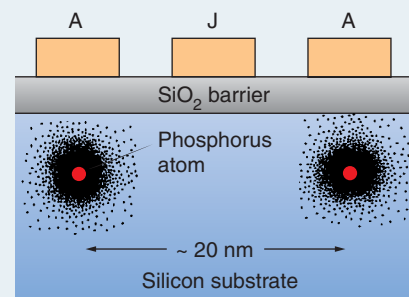


Figure 1. Kane’s Architecture for a Quantum Computer

In Kane’s concept of a silicon-based quantum computer, the qubits are phosphorus atoms embedded in an isotopically pure ^{28}Si crystal at a distance of about 20 nm from each other. Above the silicon, there is an insulating barrier of SiO_2 , and above that barrier, metallic gate electrodes. The A-gates help manipulate the individual qubits whereas the J-gates control the interaction between neighboring qubits.

from occupied energy states to unoccupied ones. To a first approximation, the tunneling current at a point on the surface is proportional to the local electron density of states (LDOS) in the sample. By measuring the tunneling current as a function of position, we can obtain an extremely localized map of the electronic structure of the sample’s surface.

The tip is attached to a piezoelectric scanning device, which moves it over the surface of the sample in a raster pattern. An image of the surface is thus obtained. In practice, we use a feedback loop to adjust the tip height and keep the tunneling current constant as the tip moves. (Scanning in this “constant-current” mode prevents the tip from crashing into protrusions, such as surface steps.) The resulting map of tip heights versus position can be used to construct an image of the surface that shows contours of constant LDOS. On many surfaces, this

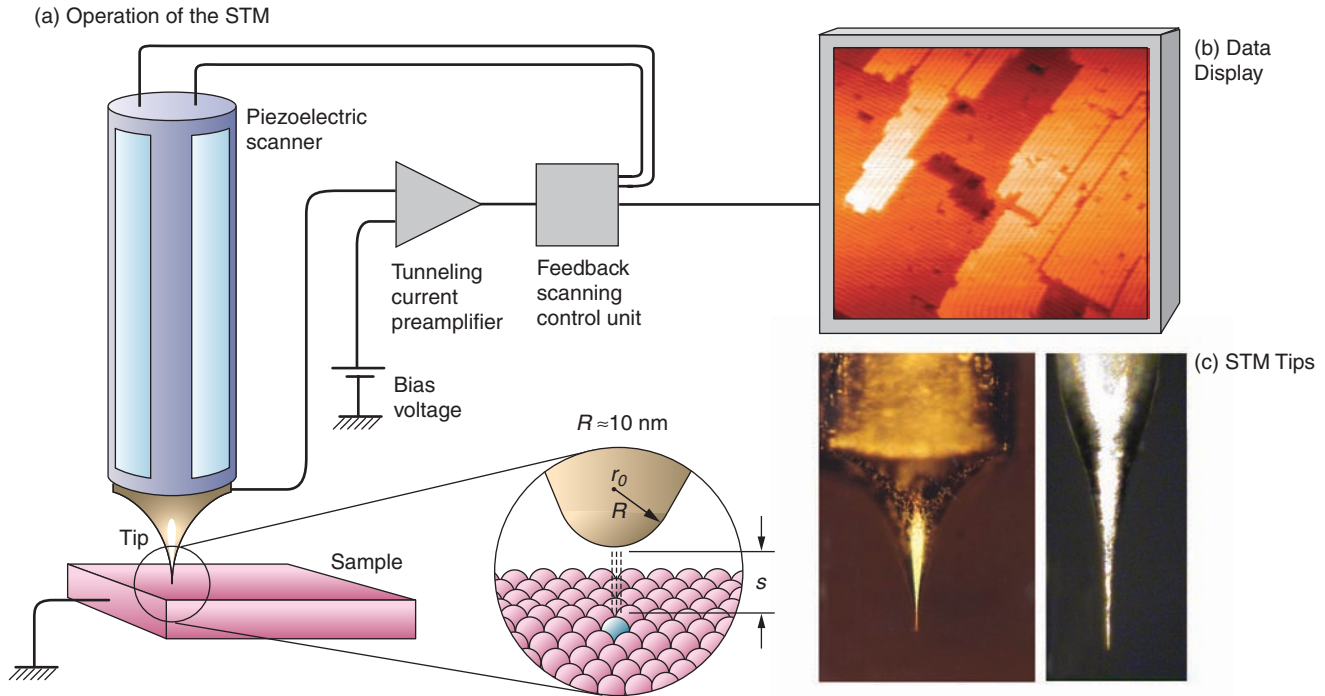


Figure 2. The Scanning Tunneling Microscope (STM)

(a) This schematic diagram illustrates the steps in the operation of an STM. An extremely sharp tip is held within a few tenths of a nanometer of a sample surface. A bias voltage V applied between the tip and the sample causes electrons to tunnel between the two. The tunneling current is monitored with a feedback loop, which keeps the current constant by varying the gap width s between the tip and the sample. The gap width is then proportional to the sample's local density

of states (LDOS). The tip moves in a raster pattern laterally over the surface. A plot of the tunneling current versus position is a map of the sample's LDOS. One such map is shown in (b). In many cases it is equivalent to a map of atomic positions. (c) These optical micrographs show two STM tips: One is made of tungsten (left) and the other, of etched 90% Pt–10% Ir alloy. Each tip has a radius of curvature of about 10 nm.

contour map is equivalent to a map of the atomic positions.

The electronic-energy diagrams of the tunneling process, shown in Figure 3, help to explain the technique's atomic resolution, as well as the subtleties of the information obtained. The applied bias voltage defines the energy offset, or energy "window," between the Fermi levels of the tip and the sample. Any electrons that have energies within that window contribute to the net tunneling current.

In 1985, shortly after the development of the STM, Jerry Tersoff and Donald Hamann described the tunneling mathematically, by applying Bardeen's tunneling theory (1961) to the tip-sample system. By assuming a

low temperature, a small bias voltage V , and a featureless tip (one in which the electron density of states is constant), they showed that the tunneling current could be written as

$$I_t \propto \sum_v |\Psi_v(r_0)|^2 \delta(eV - E_F) \quad (1)$$

Here, Ψ_v are the sample's wave functions whose energy eV above the Fermi level E_F is evaluated at the point r_0 on the tip—see Figure 2(a). The sum over the probability densities from all such wave functions is the LDOS of the sample directly below the tip, so that in the approximation of Equation (1), the tunneling current is indeed proportional to the sample's LDOS.

The spatial resolution of an STM image is extremely high (approx-

mately 0.01 angstrom) in the direction perpendicular to the surface. That is so because the tunneling probability T decreases exponentially with the separation s between the tip and the sample. The Wentzel-Kramers-Brillouin (WKB) approximation for the tunneling probability through the type of potential barrier shown in Figure 3 (a trapezoidal barrier between planar metal electrodes) yields

$$T = e^{-2\kappa s} \quad (2)$$

where κ , the inverse decay constant in the potential barrier, is given by

$$\kappa = \sqrt{\frac{2m_e}{\hbar^2} \left(\frac{\phi_t + \phi_s - eV}{2} - E \right)} \quad (3)$$

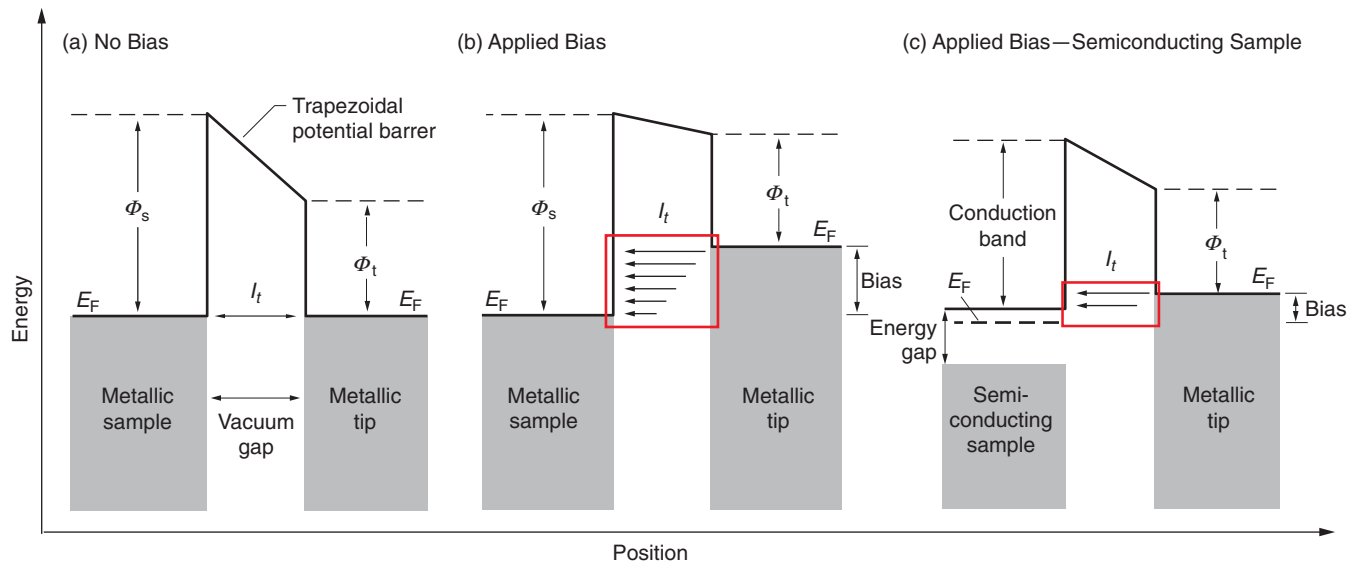


Figure 3. Tunneling Process for Metallic and Semiconducting Samples

This schematic electronic-energy level diagram helps illustrate basic STM concepts. Filled electronic states are below the Fermi level E_F whereas empty states are above. A potential barrier is created by the vacuum gap between the tip and sample. If the width of the barrier is so narrow that the electron wave functions of the tip and sample overlap, then electrons can tunnel to empty states in either the tip or the sample. (a) When the two Fermi levels are equal (because the tip and sample are connected to a common ground) there is no net current flow. (b) When a bias voltage is applied, the Fermi levels of the two materials become unequal, and the difference

defines an energy window (red box). In the case shown, the bias voltage raises the Fermi level in the metal tip relative to that in the metal sample. Electrons in filled states within the energy window can tunnel from the tip, through the potential barrier, into the sample's empty states. The arrows of decreasing size indicate that the tunneling probability is highest for electrons at the Fermi level of the tip and decreases as the electron energy decreases. (c) No states are available in the energy gap between filled states and the conduction band of a semiconductor sample. Electrons can only tunnel into empty states in the conduction band.

In Equation (3), m_e is the free-electron mass in vacuum, and \hbar is the reduced Planck constant. The variables are the work functions¹ of the tip and the sample, ϕ_t and ϕ_s , respectively, the electron kinetic energy normal to the barrier E (measured relative to the tip's Fermi level), and the bias voltage V applied to the sample.

Given nominal values for the parameters in Equation (3) (for example, $\phi_t \approx \phi_s = 3\text{--}6$ electron volts, $E \approx 0.025$ electron volts, and $V = 1\text{--}2$ volts), the decay constant κ is of the order of $0.1 \text{ nanometer}^{-1}$. A change of 0.1 nanometer in the spacing between the tip and the sample alters the tunneling probability by

¹ The work function ϕ is the energy needed to remove an electron, whose energy is at the Fermi level, from the sample.

$e^2 = 7.4$. Thus, a topographic resolution of the order of 0.001 nanometer in the direction perpendicular to the surface requires only a 2 percent precision in the measurement of the tunneling current. With carefully designed, low-noise electronics, that precision is easily achieved—even for a tunneling current of 100 picoamperes .

The resolution parallel to the surface is also atomic—on the order of 0.1 nanometer —for much the same reason: The extreme sensitivity of the tunneling current to the gap width ensures that essentially the entire tunneling current arises from a single atom or a small cluster of atoms at the very end of the tip (those atom(s) closest to the sample). On a clean, well-formed tip with a small radius of curvature, atoms or clusters that are laterally displaced from the end are

also farther from the sample and do not contribute a significant number of electrons to the tunneling current. Thus, there is very little lateral spread associated with the signal.

A more-detailed look at the origin of the tunneling current will shed additional light on the information contained in the STM image. Equation (1) can be rewritten to account for both a finite energy window for tunneling and a more-complex electronic structure of the tip as follows (Selloni 1985):

$$I_t \propto \int_{-eV}^0 \rho_t(E) \rho_s(E+eV) T(E, eV) dE, \quad (4)$$

where ρ_t and ρ_s are the tip and sample LDOS, respectively, T is the tunneling probability between the tip and the

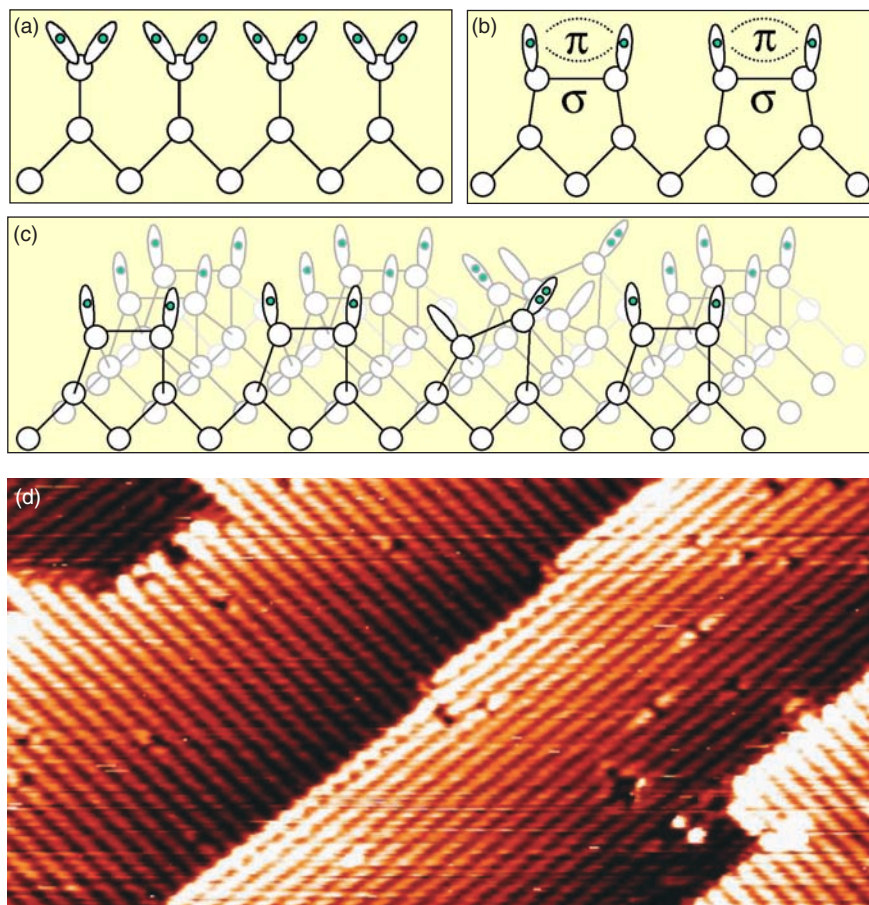


Figure 4. The Si(100)-(2 × 1) Surface

(a) Schematic view of a row of atoms in a bulk-terminated Si(100) surface. This configuration is energetically unfavorable because every atom has two singly occupied dangling bonds. (b) The bulk-terminated surface can reconstruct into the so-called Si(100)-(2 × 1) surface. The dangling bonds of neighboring atoms join to form σ -bonded dimers, and the remaining protruding bonds become weakly π -bonded. (c) This view of the reconstructed surface shows several rows of dimers. In the third row from the left, the dimers are pinned in the buckled configuration (see text). (d) A filled-state image of a Si(100)-(2 × 1) surface (10 nm × 5 nm) showing several monolayers. The bright lines making up each “terrace” are the dimer rows, which rotate by 90° with each successive layer. The defect density of an average sample is approximately 5%. The wavy line along the upper edge of the central terrace is due to buckled dimers.

sample, and the dependence of the tunneling current on r_0 has been suppressed for simplicity.

The expression in Equation (4) emphasizes that the properties of both the tip and the sample contribute to the tunneling current. Therefore, one needs to have substantial background information about both in order to interpret an STM image. For example,

numerous geometric and electronic effects go into the LDOS function ρ_s , including the electronic structure at the surface, the band structure of the material, the presence of dangling bonds at the surface or bulk resonances, the number and orientation of back bonds, and so forth. In addition, we need information about the electronic structure of any adatoms or

contaminants (such as oxygen, carbon, carbon dioxide, nitrogen, and others) that might be present. Through ρ_t , the appearance of an image is also closely related to the electronic structure of the tip. Finally, of particular importance is the fact that the electron density is not always centered about the cores of the atoms in the material. Several of these considerations arise in our work on the SSQC and will be discussed later.

One powerful technique that can be used to help us interpret images is to change the direction (sign) of the bias voltage. By doing so, we cause the tunneling current to reverse its direction. If the tip is biased to have a higher Fermi level, then current flows from the tip to the empty states in the sample. If the bias is reversed, so that the sample has a higher Fermi level, then the electrons from the sample’s filled states flow into the empty states in the tip. We therefore have a means to obtain information about the density of both the empty and filled states of the sample. The differences between the two STM images help us sort out electronic effects from structural information and to distinguish among features that appear identical when only one bias direction is used.

Despite the intricacies involved, we can interpret an STM image quite accurately when all the available information is taken into account. That is why STM imaging is continuing to produce significant results in surface science.

Preparing Silicon(100) Surfaces

The bottom-up fabrication approach begins by preparing a flat (100)-oriented silicon surface. Technologically, this is one of the most important semiconductor surfaces. For our purposes, it is relatively easy to prepare, can be patterned by

STM-based hydrogen lithography, and is well suited for the subsequent overgrowth of crystalline silicon layers. Although the (100) surface has been studied by STM and other methods for over 15 years, we are uncovering new details important to constructing the type of atomic-scale electronic structures needed in the Kane solid-state quantum computer and other quantum devices.

Figures 4(a) and 4(b) show how the bulk-terminated (100) surface, every atom of which has two dangling bonds, reconstructs in a manner that lowers the surface energy. Electrons from two neighboring silicon atoms form a σ -bond, so the resulting silicon-silicon dimer has only two dangling bonds. These bonds form a weak π -bond to further reduce the surface energy. The π -bond can easily be broken by chemically active species, such as hydrogen, which adsorb on the surface.

The reconstructed surface, commonly referred to as the Si(100)-(2 × 1) surface (a designation that derives from the corresponding electron-diffraction pattern), takes on the appearance of a neatly plowed field, with rows of dimers aligned parallel to each other, as seen in Figures 4(c) and 4(d). In filled-state STM images taken at room temperature, most of the dimers appear as symmetric bean shapes. In reality, the dimers are tilted, or buckled, and are flipping back and forth between buckled configurations very rapidly—refer to Figure 4(c). The oscillation takes place too quickly to be imaged with an STM. Therefore, in general, an average configuration is observed. Near defects or step edges, however, the dimer can be pinned in an asymmetric position and imaged. Such an image can be seen at several locations in Figure 4(d), where neighboring dimers are seen to buckle in alternate directions.

Surface preparation begins with degassing the sample and its holder

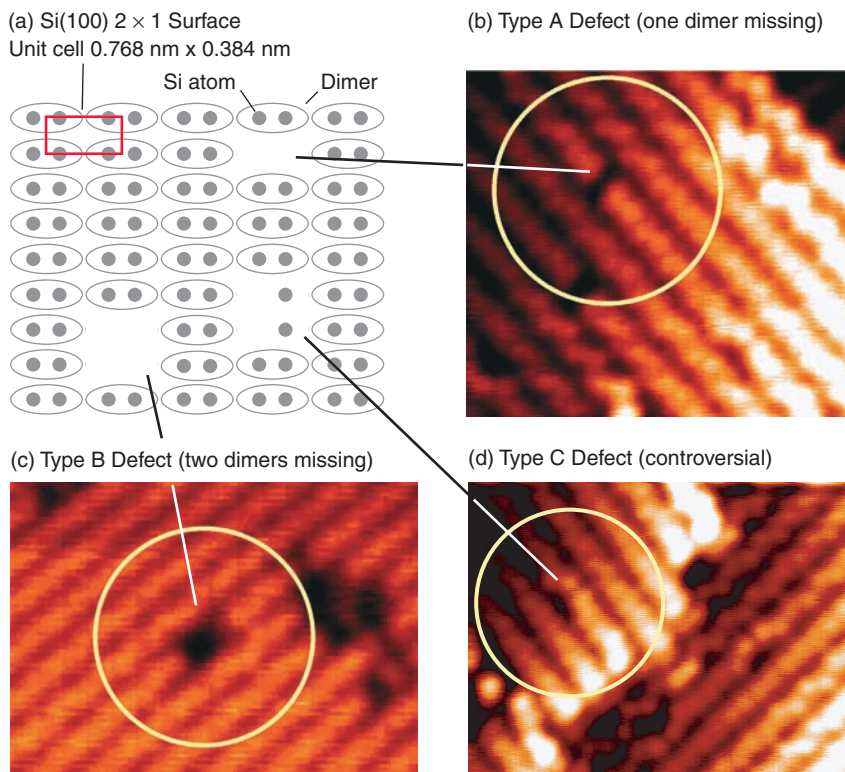


Figure 5. Defects in the Si(100)-(2 × 1) Surface

(a) This cartoon of the Si(100)-(2 × 1) surface illustrates type A, B, and C defects. (The structure of a type C defect is controversial. See text for a further discussion.) The 0.768 nm × 0.384 nm unit cell (the values are for the dimer spacings along the rows and between the rows, respectively) is also shown. (b), (c), and (d) show filled-state images of type A, B, and C defects, respectively.

by holding them at an elevated temperature for several hours. (Because the surface is reactive, this step and those that follow are carried out in ultrahigh vacuum.) The sample is flash-heated to a temperature of 1250°C and cooled under conditions that allow the surface silicon atoms to form a well-ordered Si(100)-(2 × 1) surface. But the difficulty in precisely controlling the annealing process and the inability to cut the starting substrate exactly on axis result in a surface typically consisting of several terraces of simple atomic planes. On a given terrace, all the dimer rows run in the same direction, whereas the in-plane orientation of the dimer rows rotates by 90° from one terrace to the next. The terrace edges terminate

smoothly or roughly, depending on whether the dimer rows for that terrace run parallel or perpendicular to the terrace edges, respectively.

Even a freshly prepared Si(100)-(2 × 1) surface will contain defects. The main types observed in STM images, illustrated in Figure 5, are type A defects, in which a single silicon dimer is missing, type B, in which two adjacent dimers in a row are missing, and type C, whose make-up is still controversial. Type C defects could be the result of a subsurface vacancy, or else consist of two missing silicon atoms from adjacent dimers in a row. They could also be due to an adsorbed impurity, for example an adsorbed water molecule (Chander et al. 1993). Although it is nearly impos-

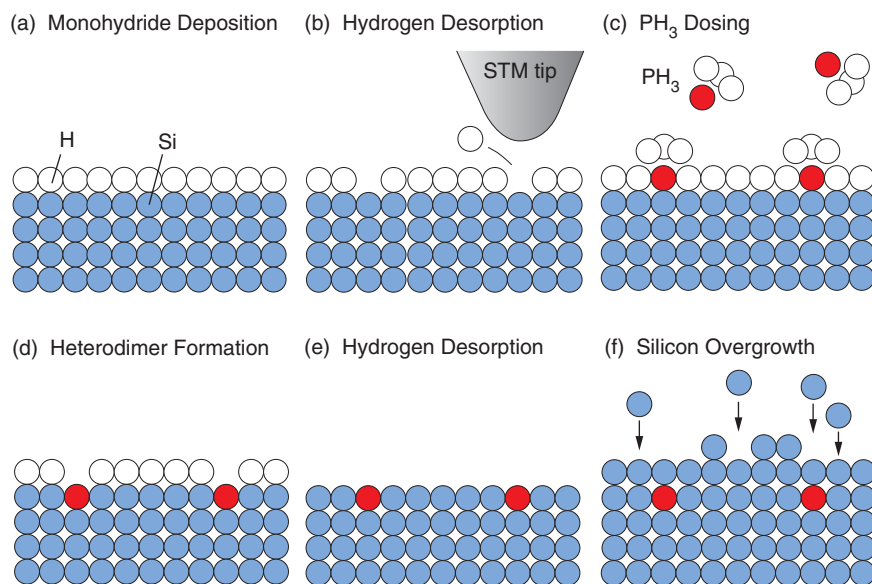


Figure 6. The Bottom-up Approach for Fabricating an Array of Phosphorus Qubits

(a) After preparing and cleaning a silicon surface, we dose it with hydrogen, which adsorbs as a monoatomic layer. (b) The STM tip selectively desorbs individual hydrogen atoms and exposes silicon at a set of regularly spaced sites that will define the qubit array. (c) PH_3 is introduced into the vacuum chamber. It bonds to the silicon only at the exposed sites. (d) A critical anneal is performed to incorporate the phosphorus atoms into the silicon surface, forming a P-Si heterodimer. (e) The hydrogen monolayer can be removed by further annealing at a slightly higher temperature (this step may not be necessary). (f) With molecular-beam epitaxy, the phosphorus array is buried under fresh layers of silicon.

sible to eliminate these defects during preparation, we can prepare surfaces with defect densities of less than a few percent by following careful vacuum practices in the STM chamber. The effect of defects on the operation of a quantum computer will be further discussed under the section “Qubits, Defects, and Dopants.”

STM-Based Hydrogen Lithography

Once we have prepared a clean surface with a low defect density, we are ready to begin the array fabrication scheme. We use a resist technology analogous to the lithographic techniques used in conventional electronics manufacturing, the main difference

being that the STM-based technology allows us to create features on the atomic scale.² The idea is illustrated in Figure 6.

The first step is to deposit a single layer of hydrogen atoms (the “resist”) on the clean surface. In order to do so, we dissociate molecular hydrogen gas by passing it over a hot filament as it enters the STM vacuum chamber. The resulting hydrogen atoms are directed onto the heated sample surface, where they break the weak π -bond and adsorb

² Scanning tunneling microscopy can be used directly to create atomically precise structures of metal atoms on metal surfaces. We are forced to adopt a lithography approach because the strong covalent bonds on the silicon surface prevent us from directly rearranging atoms using the STM.

to the surface by attaching to the very reactive dangling bonds. Provided the conditions are right, one hydrogen atom can covalently bond to each silicon atom, and the surface becomes coated with a uniform monohydride layer (see Figure 7).

The STM tip is then used as the lithographic patterning tool. Controlled-voltage pulses applied between the tip and the sample cause very small patches of the monohydride layer to vibrate and heat up and/or to become electronically excited. Individual hydrogen atoms are liberated, and as a result, the dangling bond of the underlying silicon atom becomes exposed. The tiny, atom-sized holes created by the STM are the only reactive sites on the otherwise unreactive monohydride layer.

Interestingly, the holes created in the hydrogen layer appear as protrusions above the hydrogen-terminated surface. This is an example of electronic effects influencing the STM images. Whereas the hydrogen-terminated structures protrude farther into the vacuum than the dangling bonds, the energy of the dangling bonds is closer to the window between the Fermi levels of the tip and the sample. The dangling bonds, therefore, contribute more strongly to the tunneling current and appear “taller.”

Next, we introduce high-purity phosphine (PH_3) gas directly into the ultrahigh-vacuum chamber of the microscope. The PH_3 is very reactive and adheres to the exposed dangling bond with a sticking coefficient of one. As seen in Figure 8, we can place single phosphorus-bearing molecules where necessary and thereby build an atomic-scale phosphorus array. The reacted sites appear taller than both the hydrogen-terminated sites and the unreacted dangling bonds. This effect is likely due to a combination of electronic and physical effects.

The next step is to stimulate the phosphorus atoms within the phosphine molecule (which is attached to the silicon atoms by a single bond) to

incorporate into the top layer of the silicon surface and form a phosphorus-silicon heterodimer. In that structure, the phosphorus atom takes the place of one of the silicon atoms in the dimer and attaches to the remaining silicon surface through three strong covalent bonds. Formation of the heterodimer is a critical step because it secures the phosphorus atom in its patterned location and helps prevent its diffusion during subsequent processing steps.

Before studying the mechanism for incorporation through the hydrogen resist, we had to learn how to distinguish the postdosing phosphorus-related species from other features on the silicon surface because, to date, very few reports exist on the STM imaging of single phosphine molecules on silicon. We, therefore, conducted a series of experiments in which the clean Si(100)-(2 × 1) surface was subjected to various dosing conditions. Each time, the presence of phosphorus on the surfaces was confirmed by Auger electron spectroscopy. By examining both filled- and empty-state STM images, we found it was possible to distinguish between phosphine-related surface species and surface defects.

We then faced the challenge of phosphorus incorporation. It is well known that, at room temperature, phosphine adsorbs onto a clean Si(100)-(2 × 1) surface and quickly dissociates to form PH₂ and H. Subsequent heating of the surface to about 400°C leads to the complete dissociation of PH_x (x = 2–3). We have demonstrated that, at these temperatures, the individual phosphorus atoms also incorporate into the surface and form the phosphorus-silicon heterodimer (see Figure 9). The hydrogen remains on the surface as a monohydride. Continued heating of the surface to higher temperatures will liberate the hydrogen. In this way, the surface is left clean, consisting of only silicon dimers and phosphorus-silicon heterodimers.

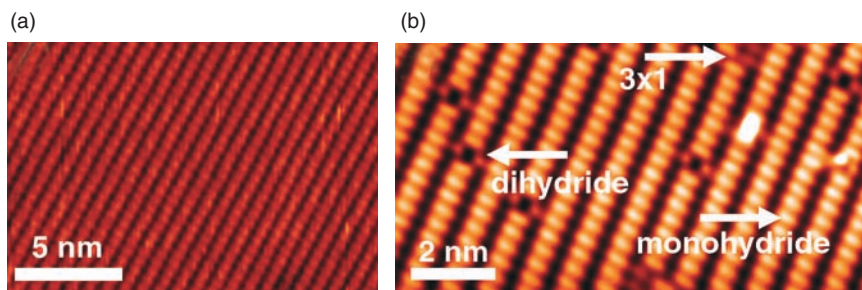
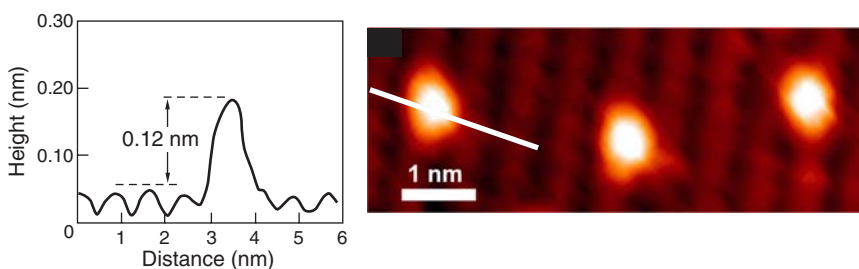


Figure 7. Creating the Hydrogen Resist

(a) This filled-state STM image is of a clean Si(100)-(2 × 1) surface with a very low defect density. (b) Shown here is a hydrogen-terminated Si(100)-(2 × 1) surface, which is almost entirely monohydride; that is, one hydrogen atom is bonded to each silicon atom. Several other structures are also apparent: dihydrides (two hydrogen atoms have bonded to a single silicon atom) and a 3 × 1 structure (three hydrogen atoms have bonded to one silicon atom).

(a) Desorption of Single Hydrogen Atoms in Monohydride Layer



(b) Adsorption of Phosphine in Desorption Sites

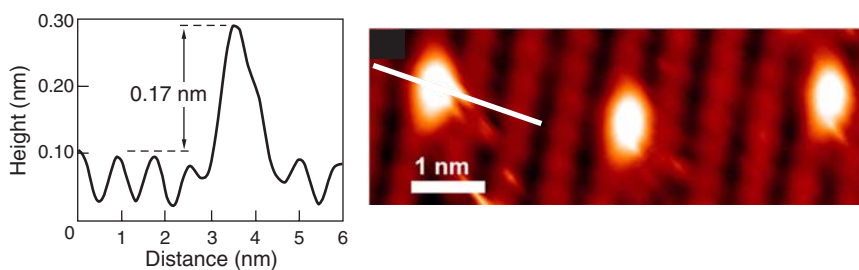


Figure 8. Adsorption of Single Phosphine Molecules

(a) This STM image (right) shows three desorption sites in a monohydride layer, and the graph (left) shows the line profile, taken along the indicated white line, of the leftmost site. The bright protrusion at each of the desorption sites is the signature of the single silicon dangling bond after desorption of just one hydrogen atom. (The sites appear brighter because their DOS are closer to the Fermi level, so they contribute more to the tunneling current.) (b) The same sites after dosing the surface with phosphine gas. The profile shows an increase of 0.05 nm in height (calibrated against an atomic step edge on the same surface), a reproducible increase that is observed at all adsorption sites. Given the information we gathered by scanning tunneling microscopy, our interpretation of the increase in height is that phosphine has adsorbed to the exposed sites.

Molecular Beam Epitaxy of Silicon

Subsequent steps for the fabrication of the SSQC call for growing a 50- to 200-angstrom-thick layer of crystalline silicon over the array of phosphorus atoms, depositing an insulating layer of SiO_2 , and aligning gate electrodes to the now buried phosphorus array. High-quality crystalline, or epitaxial, growth of silicon on silicon is typically done at high temperatures.

However, it is known that at high temperatures, phosphorus atoms buried in silicon tend to diffuse upwards and pop up to the surface. Furthermore, we observed during our incorporation studies that, at temperatures of 650°C and above, the phosphorus becomes mobile. It breaks from the heterodimer and begins to migrate about the surface until it meets another phosphorus atom. It then forms P_2 (or possibly P_4), which desorbs from the surface. Thus, the next significant question in the bottom-up approach is, “Can crystalline silicon be grown on either a clean or monohydrided surface at temperatures low enough to prevent the diffusion and segregation of phosphorus?”

Taking into account results from the literature and our own experiments, we have adopted two parallel growth strategies. We begin both by annealing the sample directly after phosphine dosing, so that the phosphorus atoms become incorporated into the silicon surface and the hydrogen resist can desorb. We then need to encapsulate the phosphorus atoms under a few monolayers of silicon. In the first growth strategy, we will grow the encapsulation layer at room temperature. The resulting layer will have a high surface roughness with numerous silicon islands and require a subsequent annealing step for surface flattening. In the second strategy, we

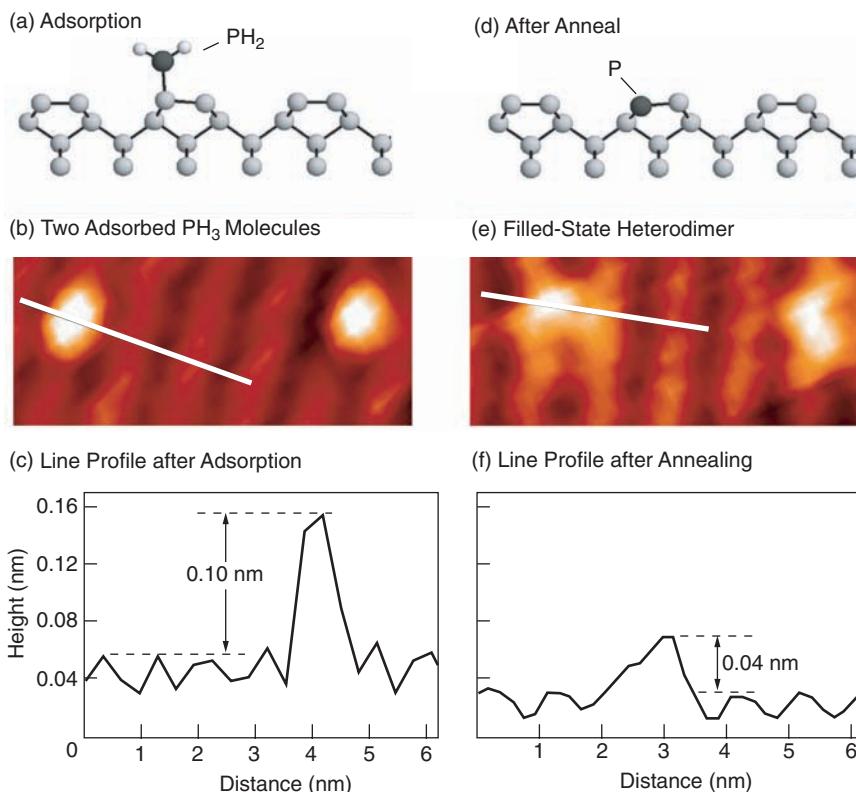


Figure 9. Incorporation of Phosphorus into the Surface

(a) This schematic diagram illustrates how phosphine molecules adsorb onto the bare $\text{Si}(100)-(2 \times 1)$ surface. The filled-state STM image in (b) is of a pair of adsorbed phosphine molecules, and (c) shows the line profile through the left molecule. (d) After annealing the surface to 400°C , the phosphorus atom incorporates into the silicon surface and forms a Si-P heterodimer. (e)–(f) These figures show the filled-state STM image of the heterodimer and the corresponding line profile. A comparison between (c) and (f) shows that there is a characteristic height difference between the nonincorporated and incorporated phosphorus, the former extending higher above the surface plane.

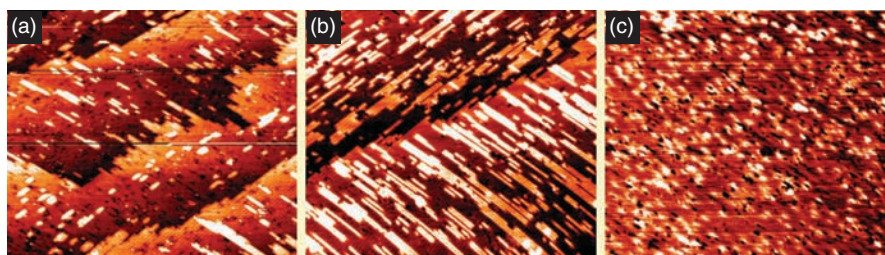


Figure 10. Images of Different Silicon Coverages

These images ($100 \text{ nm} \times 85 \text{ nm}$) of a $\text{Si}(100)-(2 \times 1)$ surface show different stages of epitaxial silicon growth. The silicon was deposited while the sample remained in the microscope and was held at about 250°C . (a) The sample is shown after a 0.08 monolayer was grown. Epitaxial growth is demonstrated by the elongated shape of the islands and their direction being perpendicular to the underlying dimer rows. (b)–(c) The sample is shown after a deposition of 0.5 monolayer and a complete monolayer, respectively. At the growth temperature noted above, the surface is rough. Defects and silicon vacancies dominate the topography.

will try to grow the silicon at an elevated temperature. Because the layers will grow epitaxially, we can eliminate the subsequent anneal, but the challenge will be to find a growth temperature that also minimizes the segregation and diffusion of the phosphorus atoms.

A significant number of experiments need to be conducted to determine the optimal encapsulation conditions. By integrating a small silicon evaporator into the STM chamber, we have already begun to study the epitaxial deposition of thin silicon layers at low temperatures. Figure 10 shows growth in the thickness of silicon of up to one monolayer at 250°C. The new layer grows epitaxially. Before it is annealed, the complete monolayer still exhibits vacancies that are not filled during the silicon overgrowth. Their possible detrimental effects on the operation of the quantum computer will have to be evaluated.

We have also begun to explore the first growth strategy (see Figure 11). We incorporated phosphorus into the silicon surface, deposited a few monolayers of silicon at room temperature, then annealed the sample for 1 minute at 250°C. As seen in Figure 11(b), this surface was fairly coarse and not suitable for subsequent epitaxial growth. A flat surface structure with island-free terraces was observed only after the sample had been annealed at 600°C. Figure 11(c), however, shows that, at those elevated temperatures, the phosphorus atoms have diffused to the surface. Although that result is disappointing, we are not discouraged. Ours are the first such studies of phosphorus encapsulation and silicon overgrowth. The preliminary results simply demand that we look for a new way to obtain a flat surface at lower annealing temperatures or an alternative way to inhibit phosphorus diffusion.

We have, however, settled the question of whether the incorporated phos-

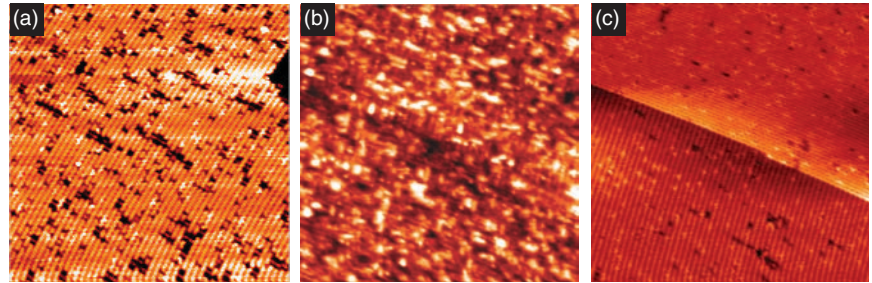


Figure 11. Silicon Overgrowth and Annealing after Low PH_3 Dosing (a) A $\text{Si}(100)-(2 \times 1)$ surface is shown after low PH_3 dosing and annealing to incorporate phosphorus atoms into the Si-P heterodimers. The heterodimers are visible as bright zigzag structures. The image size is $50 \times 50 \text{ nm}^2$. (b) The epitaxially overgrown surface is shown after annealing at 250°C. The image size is also $50 \times 50 \text{ nm}^2$. The surface is too coarse for the SSQC and must be annealed. (c) After annealing at 600°C, the surface is flat. The bright spots indicate, however, that phosphorus has diffused to the surface. The image size is $55 \times 55 \text{ nm}^2$.

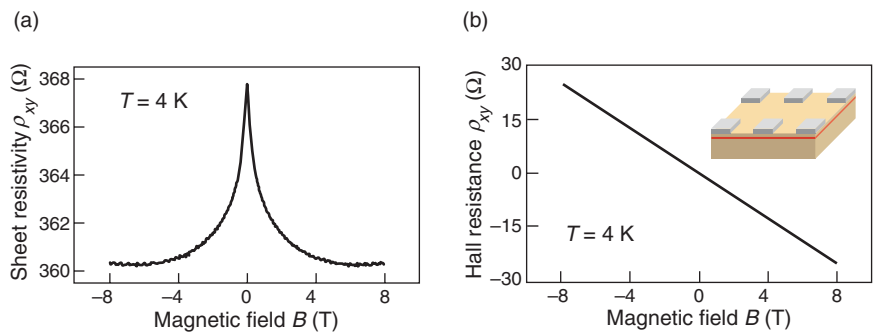


Figure 12. Electrical-Activity Tests

We wanted to check that phosphorus atoms incorporated in the silicon surface are electrically active. (a) The longitudinal resistivity ρ_{xx} of the delta-doped sample as a function of magnetic field was measured at 4 K. From this curve, a strong negative magnetoresistance is clearly observed, and it indicates the two-dimensional (2-D) nature of the delta-doped layer. (b) The Hall resistivity ρ_{xy} of the sample gives a 2-D carrier density of $2.0 \times 10^{14} \text{ cm}^{-2}$. This number agrees with our dopant density and indicates that each phosphorus dopant is electrically active. The inset is a schematic of the phosphorus delta-doped silicon sample with metal surface contacts in the van der Pauw arrangement.

phorus atoms are electrically active, that is, whether their donor electrons are free to conduct. We first grew a thin layer of phosphorus on a silicon substrate and buried it under a thick silicon layer (grown at the relatively low growth temperature of 250°C), creating a so-called delta-doped layer. According to the literature, our growth and annealing conditions resulted in a two-dimensional (2-D) density of 1.7×10^{14} phosphorus atoms per centimeter squared. If each

atom is electrically active, it would contribute one free electron to the substrate. When we measured the electron density through the Hall effect at a sample temperature of 4 kelvins, the result was a 2-D density of 2.0×10^{14} electrons per centimeter squared (see Figure 12). As the two numbers agree within measurement errors, it seems that all the phosphorus atoms are electrically active (Oberbeck et al. 2002). This result suggests that the phosphorus atoms

are incorporated in substitutional, rather than interstitial, sites, which is the ideal environment for the SSQC qubits.

Qubits, Defects, and Dopants

Although we have a clear strategy for creating and burying the phosphorus array in a working quantum computer must also be free from crystal impurities and defects. In general, defects disrupt the crystal structure and can create new pathways for quantum decoherence, which would inhibit qubit operations. Charged defects can be particularly disruptive. If the charge arises from an unpaired electron, then by necessity, there is an “impurity” spin that can interact with a qubit and affect its quantum state. Furthermore, the Coulomb potential of a charged defect lying close to a qubit can interfere with gate operations because it can offset the voltage applied to the qubit-controlling gate electrode.

Fortunately, the STM allows us to check the status of the buried qubits and charged defects during the fabrication of the quantum computer. Scanning tunneling microscopy is routinely used in characterizing the charge of individual defects found on the cleaved surfaces of compound semiconductors (Zheng et al. 1994, Lengel et al. 1994, Ebert et al. 1996). The charge becomes visible because of the so-called charge-induced band bending, illustrated in Figure 13. The states made available by band bending attract charge carriers that screen, or shield, the charged defect. Because bending shifts electronic states into or out of the window defining the source of the tunneling current, it produces a measurable enhancement or depression around the defect in the STM images. The characteristic length scale of this screening effect is given by the Debye screening length, which

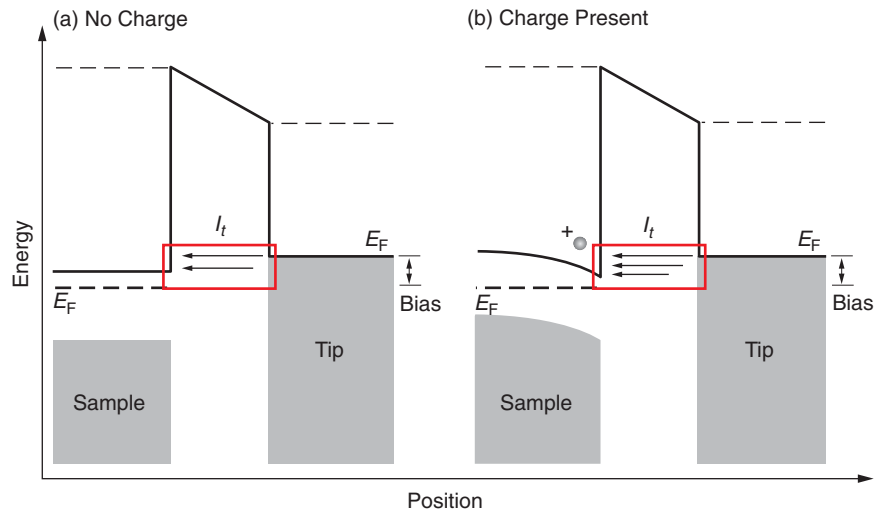


Figure 13. Band Bending

The electronic energy of an empty-state tunneling current is shown for a semiconducting sample with no charge (left) and with positive charge (right). The effects of surface states and defects have been neglected for illustration purposes. The charge-induced band bending shifts more states into the window between the Fermi levels of the tip and the sample. In this case, the increased state density relative to the rest of the neutral surface creates a long-range enhancement centered on the charge that falls off approximately like a screened Coulomb potential with a length scale set by the Debye screening length.

Table I. Expected Effect of Local Charge on Surface LDOS

Silicon(100)-2 × 1 Surface	Imaging Condition	+Charge	-Charge
<i>n</i> -type	Empty states	Enhanced	Depressed
	Filled states	No effect	Enhanced
<i>p</i> -type	Empty states	Enhanced	No effect
	Filled states	Depressed	Enhanced

depends on the semiconductor’s intrinsic properties: its dopant type and concentration (Dingle 1955).

These techniques for imaging charge have not been demonstrated on silicon surfaces until now because it has been generally assumed (based on techniques such as photoelectron spectroscopy that probe large surface areas) that the Fermi level at the surface of silicon is pinned. If that assumption is true, the bands cannot respond to charge near the surface. But by taking into account what

occurs locally and by drawing on other results obtained with the STM, we have determined that pinning of the Fermi level does not occur for clean Si(100)-(2 × 1) surfaces, except in the vicinity of type-C defects. This has allowed us to image charged defects on these clean surfaces for the first time (Brown et al. 2002).

Considering the band structure as it is currently understood, we can qualitatively determine which types of charge should be detectable in filled- and empty-state imaging on a clean

silicon(100)-(2 × 1) surface for both *p*- and *n*-type materials.³ These predictions, made under the assumptions of nondegenerate doping, a tip work function of 3 to 4 electron volts, and a low C-defect density, are compiled in Table I. As noted in the table, under some conditions, we anticipate no change in the appearance of an STM image. That result is singularly different from what is seen on compound semiconductors and arises from surface states derived from the π -bond. These states, which are not present on the compound semiconductor surfaces, limit the amount of band bending that can occur.

Based on the expectations listed in Table I, we performed STM experiments at sample biases between ± 1.5 volts on clean (2 × 1) surfaces of Si(100) samples doped with phosphorus (approximately 8×10^{15} phosphorus atoms per cubic centimeter). Low sample biases were used to ensure that effects near the band edges (for example, band bending) contributed strongly to the tunneling current. In these experiments, we were able to image charged defects consistent with our *n*-type predictions.

One such charged defect is shown in Figure 14. This defect is commonly observed in studies of Si(100)-(2 × 1) surfaces on thermally prepared samples and is typically referred to as a split-off dimer (SD, also called the 1+2 DV) defect. It consists of an A- and a B-defect on the same row, separated by one intact dimer. The empty-state image of the SD defect shows a long-range perturbation, but the filled-state image shows no corresponding feature even though the filled-state imaging is closer to the valence band edge on this *n*-type material.

These results are consistent with expectations based on Table I, indicat-

³ Electrical conduction in *n*-type materials is associated with electrons. In *p*-type materials, it is associated with holes.

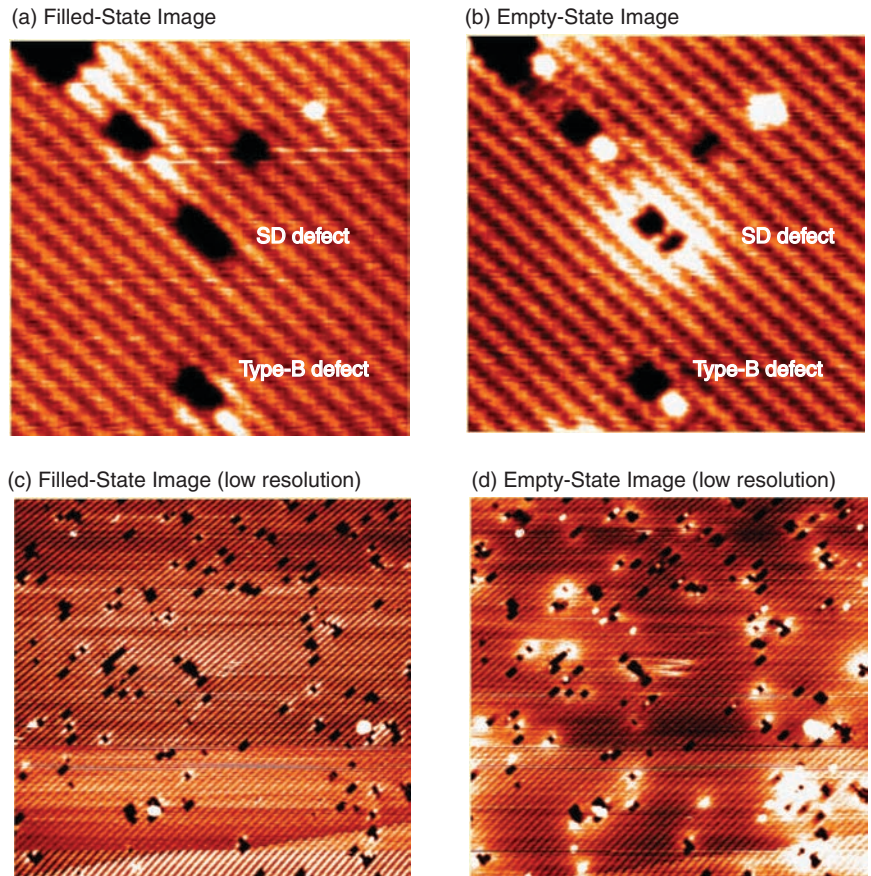


Figure 14. Finding Charged Defects

(a) This filled-state image shows a charged split-off dimer (SD) defect on a Si(100)-(2 × 1) surface (center) and a type B defect (bottom center). (b) The same defects are shown for an empty-state image. The bright “cloud” in the empty states around the central defect indicates that there is an associated positive charge. The other defects in the image appear neutral. These images measure 16.8 nm × 16.8 nm and were acquired “simultaneously” by raster scans interleaved at each bias. The asymmetric nature of the screening signature in the two biases is consistent with our expectations. (c)–(d) Pictured here are filled- and empty-state images from a different sample, taken at lower resolution. Only about one-third of the defects are charged. The images are 57 nm × 55 nm.

ing positive charge associated with the SD defect. The enhanced density of states—bright region in Figure 14(b)—appears to be nearly radially symmetric and approximately centered on the defect structure. Sections through the data show that the signature is discernible out to about 4.5 nanometers from the center. The corresponding Debye screening length, obtained from a screened Coulomb potential function fit to the

sections, is approximately 3 nanometers. That result was unexpected. The typical bulk value for the screening length that is consistent with our dopant density (which correlates with the number of charge carriers) is several tens of nanometers. The short screening length indicates a high dopant density at the surface. One explanation is that, on our thermally prepared surface, buried dopants may diffuse because of the high

temperature, and the actual density at the surface could be high enough to account for the 3-nanometer screening length.

Another interesting result is that, for samples with defect densities less than 5 percent, only about one-third of the surface defects are charged. This finding tells us that the charge is not associated with the simple vacancy structure observed in the images but must arise from more subtle effects. Charged defects may be due to, for example, rebonding differences among second-layer atoms. Charged and neutral defects may also coexist because of subsurface impurities or gas-phase species adsorbed in the vacancy structure itself. At this point and by using only scanning tunneling microscopy, we are unable to ascertain why only some defects are charged.

The fact that charge can be imaged on a silicon surface tells us that, after creating a flat overlayer, we will be able to detect the subsurface charged qubits. This finding is important for determining whether the qubits move during subsequent silicon growth. And looking beyond the Kane architecture, our results will be applicable to any implementation of a solid-state, silicon-based quantum computer.

Future Challenges

To date, we have demonstrated most of the individual steps required to successfully fabricate the Kane SSQC. We can create a small phosphorus array (O'Brien et al. 2001) and incorporate that array into the silicon surface. We have shown that the phosphorus atoms remain electrically active (Oberbeck et al. 2002). We can grow silicon epitaxially in the STM at a temperature that should leave the array intact, and we can detect charged defects at the surface. Although not reported in this article,

the Semiconductor Nanofabrication Facility housed at the University of New South Wales in Sydney, Australia, has fabricated metallic gates with dimensions close to those required for proper operation of the quantum computer.

As we integrate the aforementioned steps and try to produce a few-qubit device, several questions remain to be answered. Will the qubit array stay intact during silicon overgrowth and during any required postanneals? Can we remove defects during fabrication and, if not, to what extent will vacancies or impurities affect the computer operation? Will we introduce charge defects at the interface between the silicon overlayer and the insulating layer? How well can we register the gates with the qubits, once the array has been built?

Still, the number of questions that confront us today is far smaller than the number that faced us three years ago, when we first contemplated the steps involved in fabricating the SSQC. At that time, each question was tied to a long list of experimental obstacles that needed to be overcome. Through the combined efforts of two laboratories in the United States and Australia, we have been able to develop experimental procedures that have moved us closer to fabricating a qubit array. Given our prior success, we are hopeful that the remaining issues can be addressed successfully as well.

On a different note, one exciting idea that has emerged recently is the possibility that STM can detect single spins. Yshay Manassen et al. (2000) reported detection of a spin-induced alternating-current component in the STM tunneling current. Recent theoretical work, discussed in the article "Theory of Single-Spin Detection with a Scanning Tunneling Microscope" on page 184, offers an explanation and puts the experimental finding on firmer ground. At

Los Alamos, we are in the process of modifying the electronics of our STM and adding an external magnetic field with the hope of confirming the effect. If we are successful, directly studying spin-spin interactions and creating, manipulating, and reading out surface-bound qubits may become reality. Such a possibility is indeed exciting. ■

Acknowledgments

This work would not have been possible without the effort of many students and postdoctoral researchers. They include Jeremy O'Brien, Steven Schofield, Neil Curson, Lars Oberbeck, and Holger Grube.

Further Reading

- Bardeen, J. 1961. Tunneling from a Many-Particle Point of View. *Phys. Rev. Lett.* **6**: 57.
- Brown, G. W., H. Grube, M. E. Hawley, S. R. Schofield, N. J. Curson, M. Y. Simmons, and R. G. Clark. 2002. Imaging Charged Defects on Clean Si(100)-(2 × 1) with Scanning Tunneling Microscopy. *J. Appl. Phys.* **92**: 820.
- Chander, M., Y. Z. Li, J. C. Patrin, and J. H. Weaver. 1993. Si(100)-(2 × 1) Surface Defects and Dissociative and Nondissociative Adsorption of H₂O Studied with Scanning Tunneling Microscopy. *Phys. Rev. B* **48**: 2493.
- Chen, C. J. 1990. Origin of Atomic Resolution on Metal Surfaces in Scanning Tunneling Microscopy. *Phys. Rev. Lett.* **65**: 448.
- Dingle, R. B., 1955. *Phil. Mag.* **46**: 831.
- Ebert, P., X. Chen, M. Heinrich, M. Simon, K. Urban, and M. G. Lagally. 1996. Direct Determination of the Interaction between Vacancies on InP(110) Surfaces. *Phys. Rev. Lett.* **76**: 2089.
- Ebert, P., M. Heinrich, M. Simon, C. Domke, K. Urban, C. K. Shih et al. 1996. Thermal Formation of Zn-Dopant-Vacancy Defect Complexes on InP(110) Surfaces. *Phys. Rev. B* **53**: 4580.
- Kane, B. E. 1998. A Silicon-Based Nuclear Spin Quantum Computer. *Nature* **393**: 133.
- Lengel, G., R. Wilkins, G. Brown, M. Weimer, J. Gryko, and R. E. Allen. 1994. Geometry

- and Electronic Structure of the Arsenic Vacancy on GaAs(110). *Phys. Rev. Lett.* **72**: 836.
- Liu, L., J. Yu, and J. W. Lyding. 2001. Atom-Resolved Three-Dimensional Mapping of Boron Dopants in Si(100) by Scanning Tunneling Microscopy. *Appl. Phys. Lett.* **78**: 386.
- Lyding, K. W., T.-C. Shen, J. S. Hubacek, J. R. Tucker, and G. C. Abeln. 1994. Nanoscale Patterning and Oxidation of H-Passivated Si(100)-(2 × 1) Surfaces with an Ultrahigh Vacuum Scanning Tunneling Microscope. *Appl. Phys. Lett.* **64**: 2010.
- Manassen, Y., I. Mukhopadhyay, and N. R. Rao. 2000. Electron-Spin-Resonance STM on Iron Atoms in Silicon. *Phys. Rev. B* **61**: 16223.
- McEllistrem, M., G. Haase, D. Chen, and R. J. Hamers. 1993. Electrostatic Sample-Tip Interactions in the Scanning Tunneling Microscope. *Phys. Rev. Lett.* **70**: 2471.
- Oberbeck, L., N. J. Curson, M. Y. Simmons, R. Brenner, A. R. Hamilton, S. R. Schofield, and R. G. Clark. 2002. Encapsulation of Phosphorus Dopants in Silicon for the Fabrication of a Quantum Computer. (To be published in *Appl. Phys. Lett.*)
- O'Brien, J. L., S. R. Schofield, M. Y. Simmons, R. G. Clark, A. S. Dzurak, N. J. Curson et al. 2001. Towards the Fabrication of Phosphorus Qubits for a Silicon Quantum Computer. *Phys. Rev. B* **64**: 161401(R).
- Owen, J. H. G., D. R. Bowler, C. M. Goringe, K. Miki, and G. A. D. Briggs. 1995. Identification of the Si(001) Missing Dimer Defect Structure by Low Bias Voltage STM and LDA Modeling. *Surf. Sci.* **341**: 11042.
- Selloni, A., P. Carnevali, E. Tosatti, C. D. Chen. 1985. Voltage-Dependent Scanning-Tunneling Microscopy of a Crystal Surface: Graphite. *Phys. Rev. B* **31**: 2602.
- Tersoff, J., and D. R. Hamann. 1985. Theory of the Scanning Tunneling Microscope. *Phys. Rev. B* **31**: 805.
- Ukraitsev, V. A., Z. Dohnalek, and J. T. Yates, Jr. 1997. Electronic Characterization of Defect Sites on Si(001)-(2 × 1) by STM. *Surf. Sci.* **388**: 132.
- Yates, J. T. Jr., and V. A. Ukraitsev. 1996. The Role of Nickel in Si(001) Roughening. *Surf. Sci.* **346**: 31.
- Zheng, J. F., X. Liu, N. Newman, E. R. Weber, D. F. Olgetree, and M. Salmeron. 1994. Scanning Tunneling Microscopy Studies of Si Donors (Si_{Ga}) in GaAs. *Phys. Rev. Lett.* **72**: 1490.
- Marilyn E. Hawley** received her Ph.D. in physics from The Johns Hopkins University in 1987. Marilyn joined Los Alamos National Laboratory as a postdoctoral fellow in 1989. She was the first to successfully image and identify the spiral growth mechanism in high-temperature superconducting films using scanning tunneling microscopy (STM), results that were featured on one of the covers of *Science*. Marilyn has over 18 years of experience in STM and many advanced atomic force microscopy techniques. In 1991, she became a staff member in the Center for Materials Science at Los Alamos. Later, Marilyn established the Scanning Probe Microscopy Laboratory, a facility devoted to the development and use of scanning probe techniques for research on a broad spectrum of materials. She currently leads the Los Alamos STM team in an effort aimed at fabricating a qubit array for a possible future solid-state quantum computer.



- Geoff Brown** received a bachelor's degree in physics from Abilene Christian University and a Ph.D. in physics from Texas A&M University, where he studied the physics and chemistry of semiconductor surfaces using scanning tunneling microscopy (STM). In 1996, Geoff started at Los Alamos as a postdoctoral fellow in the Center for Materials Science, and he is now a technical staff member in the Scanning Probe Microscopy Laboratory at Los Alamos. His research involves STM and atomic-force microscopy studies of a wide range of materials, including semiconductors, complex oxides, metals, and superconductors.



- Michelle Y. Simmons** is currently the Director of the Atomic Fabrication Facility and a Queen Elizabeth II Research Fellow at the University of New South Wales, in Sydney, Australia. She joined the Centre for Quantum Computer Technology as the Program Manager in Atomic Fabrication and Crystal Growth in 2000, after completing a postdoctoral fellowship at the University of Cambridge, in the United Kingdom, where she was in charge of the design, fabrication, and characterization of ultrahigh-quality quantum electronic devices. She has over 14 years of experience in all aspects of semiconductor crystal growth, device fabrication, and electrical characterization of quantum electronic devices. Her current research interests are to understand how quantum electronic devices work as they become purer and smaller and to use this knowledge toward building the next generation of devices by using quantum principles—in particular a silicon-based quantum computer.



For the biography of Robert Clark, see page 300.