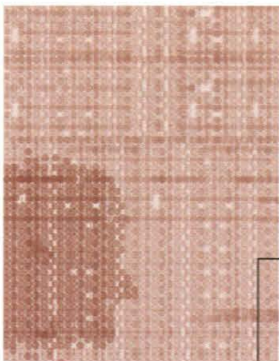
CURJ

NOVEMBER 2002
VOL.2 NO.2

CALTECH UNDERGRADUATE RESEARCH JOURNAL

PRIVACY THROUGH UNCERTAINTY

CURJ



LETTERS

REVIEWS

RESEARCH

Ap *Applied Physics*
B *Biology*
Cb *Computational Biology*
Ce *Chemical Engineering*
Cs *Computer Science*
P *Physics*

FINIS

FROM THE EDITOR

Only months after the emotional commemoration of September Eleventh, talk of war once again casts a long shadow over America. Political speculations about preemptive strikes can be found daily in newspaper headlines. The focus on national security diminishes the attention to fostering economic growth, providing better healthcare, and improving education.

By bringing the war on terrorism to the doorsteps of the few that plan these atrocious acts, our methods of warfare are becoming increasingly similar to what is now accepted as a fact of life for millions in the Middle East. As the face of combat changes, the price for flushing out terrorists increases. In Afghanistan, servicemen tell reporters that in the endless search for terrorists, the daily intrusions into civilian homes have significantly interrupted civilian lives. Innocent lives are now counted as casualties due to cultural misunderstandings and malfunctioning precision bombs.

The global crackdown on terrorism has been fueled by America's search for justice. Yet in this call for justice, the loss of civil rights is being justified. New European laws have been implemented to detain foreigners. The United States has opened lanes of communication and trade with Arab regimes, some closed for years due to impasses reached regarding democratic development, in exchange for suspect lists and additional information on alleged terrorists.

In this war where the enemy no longer occupies trenches across a battlefield, as a nation we must be reminded of our limits. We must ask ourselves what we are willing to sacrifice in order to prevent terrorist acts. If the answer lies in the development of a Big Brother state, will we be willing to compromise the freedoms that currently define our society? Let us hope that if, as a nation, we are to choose between forgoing freedoms and attempting to prevent catastrophe, history will remember us as those who recognized our limits rather than those who compromised our fundamental ideals.

Philip Wong
Editor-in-Chief

CURJ

EXECUTIVE EDITORS

Editor-in-Chief
Philip Wong

Executive Content Editor
Shaun P. Lee

Art Director/Designer
Hasnain S. Khimji

Executive Online Editor
Loban A. Rahman

Operations Manager
Sindy Tang

ASSOCIATE EDITORS

Associate Content Editors
Jordan Boyd-Graber
Wei Lien Dang
Sarah Wilhoit

Associate Online Editors
Vivek Chandran
Robert Li
Timothy Wan

Associate Print Editor
Michelle Giron

Associate Operations Manager
Tina Hsu

STAFF COLLABORATORS

Ramone Muñoz
Graphics Advisor, Art Center College of Design

Carolyn Merkel
Director, Student-Faculty Programs, Caltech

Gillian Pierce
Science Writing Requirement Coordinator, Caltech

COVER IMAGE

Drops of Eve Hasnain S. Khimji

CURJ DONORS

Caltech Biology Division
Caltech Chemistry and Chemical Engineering Division
Caltech Engineering and Applied Sciences Division
Caltech Humanities and Social Sciences Division
Caltech President's Office
Typecraft, Inc.

CALTECH UNDERGRADUATE RESEARCH JOURNAL VOL.II NO.II

Original CURJ Design by Aniko Hullner Grau

www.curj.caltech.edu

ArtCenter College of Design

THE FOCUS OF OUR NATION

BY U.S. SENATOR BARBARA BOXER

The current session in Congress has been dominated by national defense issues, foreign policy, and domestic security due to the horrific acts that were committed one year ago on September 11. However, the current economic situation has put many Americans out of work, health care costs are on the rise, and retirement funds and stock portfolios have been hit. The outlook for students now in college hasn't been this bleak for decades: higher tuition, higher loan debt and a tough job market after graduation. To me, homeland security also means economic security for the American people. And the Bush Administration is AWOL when it comes to addressing the economy. By some measures, this is the worst economic record of any presidential administration in 50 years. Compared to previous administrations, the losses in the stock market are the worst in 50 years; the number of jobs lost in the private sector is the worst in 50 years; and the rate of growth is the worst in 50 years. People are hurting and their hopes and dreams for the future have been jolted as they watch their IRA and 401(k) holdings dwindle or, in some cases, disappear entirely.

With corporate scandals continuing to be exposed almost daily on the front pages of our newspapers, it is clear that our government should be prosecuting those responsible, stepping up its regulatory vigilance and taking other steps to prevent future similar abuses. Unfortunately, despite the best efforts of reformers in Congress, while one significant bill passed and was signed into law due to the leadership of Senator Sarbanes, others were defeated because of Administration opposition or indifference and because of lobbying by the same corporate interests responsible for the climate that fostered the scandals. And while the Sarbanes bill was a victory, the Administration is already undermining it by refusing to give the Securities and Exchange Commission the resources the legislation calls for to make it an effective corporate watchdog.

These corporate scandals have seriously undermined faith in American business and have become a factor in our depressed economy. People have lost their savings and their jobs, and in the case of our state of California, our entire state has paid a major price. Just over one year ago, California was suffering from an electricity crisis caused by corporate greed. Between 1999 and 2000, demand went up four percent while at the same time the amount the state paid for electricity increased by 266 percent. Not only did prices go up, but our people and businesses were faced with rolling blackouts. The State was forced to take over electricity purchasing, under enormous budgetary pressure, contributing to the State going from a $12 billion budget surplus to a $23.6 billion budget deficit. This situation forced the State to impose broad cutbacks and to cut funding even in priority areas such as education and health care.

The evidence continues to mount that our California energy crisis was in large part fostered by greedy energy company executives who invented scams that allowed energy prices in California to be manipulated; scams that created phony energy shortages in

California, sending the cost of electricity into the stratosphere; scams that fleeced ratepayers and our state government and caused blackouts that endangered the health, safety and security of millions of Californians.

The financially-shaky Enron bled California to prop up its profits and keep the price of its stock high so that insiders could cash out. And then Enron used its influence with the Bush Administration to ensure that it could continue to manipulate the California market. Vice President Cheney continually told us that Californians created the problem and Californians should fix it or live with it. The one entity that could have helped California–the Federal Energy Regulatory Commission (FERC)–for months did nothing to help California. For justice to be done, indictments must be handed down, refunds must be ordered, long-term contracts must be renegotiated, and cost-based pricing must remain in effect. I will continue to push for Congress to learn the truth about Enron and other companies who hurt the people I represent.

When our economy is weak and our state and local budgets are tight, one area that inevitably suffers is education. This is particularly tragic, because high quality education is the key to a successful future for young Americans and essential for a thriving economy. I am a first generation American on my mother's side and my mother never even graduated from high school because she had to work to help support her family. So every time I walk into the Senate chamber I know why I am there–to ensure that the American dream is there for all of those who work for it and play by the rules and to go after those who break the rules and take advantage of hard-working Americans.

The War on terrorism, the threat of weapons of mass destruction and national security issues are priorities and should continue to be priorities for our nation. However, they shouldn't be the only priorities. A strong economy is crucial, not only because it provides good jobs for our people, but because it is also essential for our national strength and security. We must restore confidence in American businesses by continuing to take the steps necessary to hold accountable those who are responsible and to strengthen our oversight system to prevent future abuses. I will continue to fight, not only on the broader corporate reform front, but for justice for California. We are a great nation, and I firmly believe that we must deal with urgent international security issues. At the same time we need to understand that we will not have true homeland security unless we rebuild our economy and assure quality, affordable educational opportunities, followed by good jobs, for this generation of students.

Senator Barbara Boxer is currently serving her second term, representing the State of California. She is a member of the Commerce, Science, and Transportation Committee, the Environment and Public Works Committee, and the Foreign Relations Committee.

CAN THE AMERICAN HOMELAND BE DEFENDED AGAINST CATASTROPHIC TERRORISM?

BY LEWIS M. BRANSCOMB

The best–indeed the only–effective defense is a major change in America's educational, foreign, and national security policies. It will take a reduction in American ignorance of foreign cultures and religions, a willingness to spend more on making the world livable, peaceful, and democratic than we spend on war, and a dramatic reduction in "superpower" arrogance, which ignores the advice of our friends abroad. Since this seems highly unlikely in the current political climate, Americans will have to pay attention to more practical means, under our own control, for using science and technology to make the terrorists' jobs a lot harder.

A massive study by 119 experts assembled by the three National Academies–of Science, of Engineering, and the Institute of Medicine–entitled Making Our Nation Safer: the Role of Science and Technology in Countering Terrorism identified 134 ways that technology can make the nation safer. This project was funded by the Academies from their own funds, since no government agency had both the mission and the money to ask for this work.

What will it take to allow the nation's extraordinary technical capability to address the vulnerability to terrorism? Even today the bills to create a Department of Homeland Security languish in the Senate. The administration's proposed R&D budget for counterterrorism in the new department, for the year that began on October 1, was about $500 million, or about 1.3 percent of the new department's budget, and a vastly smaller fraction of the R&D for the weapons that would be used in a war in Iraq. But more important than the lack of priority given to the science and technology for countering terrorism is the lack of both organization and policy to allow technical assets to be effectively used in this radically new situation. There are five major issues that remain to be addressed by the government.

• The federal government is organized for cold war, not for terrorism. The new security threat is both a military and a criminal threat; a threat that is both foreign and domestic. It is immediate but it will never disappear. Responsibility for protecting the public requires both government and private sector action. Only in the case of the separate jurisdictions of the FBI and the CIA has the Congress even begun to discuss how an integrated, national capability for preventing terror attacks can be achieved.

• The agencies that are most likely to comprise the new Homeland Security Depart-ment share one characteristic: none of them have broad and successful experience in the conduct of research, the development of new tools, and acquisition of complex systems. Technology and systems experience reside outside the homeland security agencies, for example in the Department of Defense. The missions and money that these agencies must have depends on the ability of the White House Office of Homeland Security (OHS) , supported by the Office of Science and Technology Policy (OSTP) to advise the Office of Management and Budget on the programs and budgets required. Both the OHS and the OSTP have small staffs and minimal budgets. Whether or not the functions and authorities of the OHS should be established by statute or not, it–and

OSTP–must have more resources and more authority to get this job done.

• Some 85 percent of the nation's energy, information, transport, and financial services infrastructure is owned and operated by private business. The cities–where the people are the target–are especially dependant on this critical infrastructure. Each of these systems is not only highly vulnerable itself, but is also connected to one or more of the others. Who will pay for the changes required to harden critical infrastructure against attack? Most of the firms are waiting for the government to decide what policy tools it will use. Among the possibilities are new regulations, co-development subsidies, and anti-trust relief to encourage voluntary collaboration. There does not seem to be any place in government where these issues are seriously and comprehensively addressed.

• How will priorities for science and technology programs be set, programs managed and tested for effectiveness? The Academies report urges government to create a technical support organization, called the Homeland Security Institute, somewhat like the dedicated, not-for-profit agencies that support decision making in the military, such as MITRE, Aerospace, and Project Airforce at RAND. It would be staffed with experts experienced in complex systems analysis, modeling and simulation, and project management. It would help the department set priorities for its science and technology strategy, help define the systems that need development, and test the effectiveness of those systems.

• Since the vulnerabilities to terrorism were not created by the terrorists but are inherent in any highly competitive economy in which every decision seeks to maximize efficiency, the vulnerability of society will never go away. New technology can also minimize impacts on civil society while protecting people and infrastructure. This suggests that a high priority should be accorded to long range, imaginative research to find new solutions. This basic research strategy has been called "Jeffersonian Science." Funds for the basic research supporting agencies (such as NSF, NIH, DOE, and NASA) would be made available specifically for exploring new ideas that might be helpful to homeland security. The agencies would identify the disciplines they believe have the best promise of coming up with good new ideas, and the research community would come up with the ideas and competitive proposals for their support.

This is a set of very big challenges for public policy, and I have not addressed a second set that are equally important–policy at the municipal, county, and state level and their relationship to federal activities. A year has gone by with very little progress at the federal level. The prospect of an attack on Iraq is diverting attention from these issues and the resources they require.

Lewis M. Branscomb is Professor, emeritus, and former director of the Science, Technology and Public Policy Program at Harvard University's Kennedy School of Government. He is a member of all three National Academies: Science, Engineering and the Institute of Medicine, and co-chaired (with Richard Klausner) the Academies' project on Science and Technology for Countering Terrorism.

CLEAN FUELS POWER THE FUTURE

BY MICHELLE K. ALLIS

IMAGINE AN EERILY QUIET METROPOLITAN FREEWAY, FREE FROM THE LOW rumble of diesel trucks, the syncopated rattling of motorcycles, and the persistent idle of passenger vehicles. Imagine major cities several degrees cooler and almost free of unhealthy smog. Fuel cells promise to make this science fiction soon a reality.

In a world where energy demands are increasing while traditional energy sources are diminishing, fuel cells can solve the problems of air and noise pollution while providing an efficient, reliable source of power. By converting the chemical energy of fuels such as hydrogen or methanol directly into electrical energy, fuel cells promise to combine the operational cleanliness of a battery with the convenience of a combustion engine. Scientists have recently made vast improvements on existing technology, improvements which could lead to the widespread adoption of various types of fuel cells within the next decade. By providing clean, reliable energy, fuel cells will alleviate concerns over unstable gas prices, rolling blackouts, and nonrenewable energy sources.

A LONG KNOWN EFFECT

Invented by Sir William Grove in 1839, fuel cells were overshadowed by inexpensive fossil fuels until the 20th century, when the National Aeronautics and Space Administration (NASA) chose to use the technology aboard the Gemini, Apollo, and shuttle missions. NASA engineers improved the reliability of fuel cells, bringing effective operational life times from hundreds of hours to thousands of hours and providing astronauts with a practical, light-weight source of electricity, heat, and water.

Fuel cells take advantage of the fact that oxygen and fuels such as hydrogen or methanol have high chemical energies. When combined under certain conditions, these reactants form lower energy products such as water and carbon dioxide. This is a favorable reaction that is driven because all chemical systems prefer to minimize their internal energies. The difference in energy between the reactants and the products is the theoretical maximum amount of energy released by the reaction that can do useful work.However, if the two chemicals are physically separated, it is impossible for them to react, no matter how chemically favorable the reaction may be. A fuel cell harnesses the potential energy of this system by preventing the oxygen and the fuel from reacting directly.

A fuel cell is composed of two main sections known as an anode and a cathode, which hold the fuel and the oxidant respectively. In most cells, a semipermeable membrane separates the two sides, allowing charged molecules to pass while blocking electrons and the electrically neutral oxidant and fuel molecules. A fuel cell gives the system a way to reach its preferred minimum energy state, but forces the system to convert the molecules into ions before they can react.

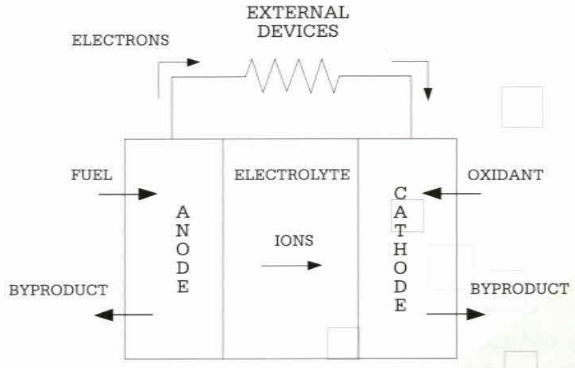


FIGURE 1. In a fuel cell, ions pass through the semipermeable electrolyte membrane, while electrons travel along the wire, producing current to power external devices.

In the simplest fuel cell, the cathode contains oxygen from the air, and the anode contains hydrogen. A catalyst such as platinum is used at the anode to increase the rate at which hydrogen gas loses its electrons to form hydrogen ions (protons) and at the cathode to increase the rate at which oxygen, protons, and electrons form water. The membrane only allows the protons to diffuse to the other side to combine with the hydroxide ions to form water. To keep the charge balanced across the membrane, the electrons must flow from the anode to the cathode by passing through an external wire. The current passing through this wire can be used to power an external electrical device such as a motor.

> "Fuel cells promise to combine the operational cleanliness of a battery with the convenience of a combustion engine."

"Because a fuel cell-powered car does not rely on **miniature explosions**, it can also operate at far **lower temperatures** and produce cleaner energy with fewer moving parts."

Other fuel cell systems can use more easily obtainable fuels such as methanol, natural gas, and even gasoline. These systems reform these hydrocarbon fuels into a hydrogen-rich gas using complex fuel processors. However, these systems add weight and these conversions must be done at high temperatures: 200 degrees Celsius for methanol, and over 800 degrees Celsius for gasoline. Moreover, the carbon and nitrogen in the fuels is produced as "waste" nitrogen gas and carbon dioxide.

Similar to an internal combustion engine, a fuel cell uses external fuel and oxygen to produce power, only at a much higher efficiency. In an internal combustion engine, the chemical energy is first converted to thermal energy by burning the fuel with oxygen. The thermal energy created by the resulting explosions is then converted to mechanical energy to power the car. However, even with great technological advances, combustion engines will never be able to convert all of the heat energy into mechanical energy.

A fuel cell converts the chemical energies of the fuel and the oxidant directly into electrical energy. This electrical energy is then used to power an electric motor, yielding fuel efficiencies two to three times higher than that of present day combustion engines. Because a fuel cell-powered car does not rely on miniature explosions, it can also operate at far lower temperatures and produce cleaner energy using fewer moving parts.

CURRENT TECHNOLOGICAL CHALLENGES

There are currently many types of fuel cells, each of which suits a different range of purposes. These types are characterized by their electrolytes, which control both the type of ions involved in the diffusion process as well as the conditions under which the fuel cell can operate. The electrolyte in a polymer eletrolyte membrane (PEM) fuel cell must stay wet to conduct protons and therefore can only operate at temperatures below 100 degrees Celsius. In contrast, solid oxide fuel cells (SOFCs) operate at temperatures between 600 and 1000 degrees Celsius, due to the high thermal energy required to mobilize the oxide ions.

The different operating temperatures of the various types of fuel cells offer certain advantages and disadvantages. An SOFC is far more efficient than a PEM fuel cell because its higher temperatures allow faster rates of reaction, but takes longer to start than a PEM fuel cell, which can operate at temperatures as low as 60 degrees Celsius. Both temperature extremes come with high costs, since low temperature fuel cells require expensive catalysts such as platinum to operate efficiently, while high temperature fuel cells must be designed and built to survive their operating conditions. The platinum catalysts in PEM fuel cells are also susceptible to poisoning by carbon monoxide, which is produced by the fuel reforming process. This limits the types of fuel a PEM fuel cell can use.

Due to the benefits and problems of high and low temperature fuel cells, there is an incentive to design a fuel cell that can operate at an intermediate temperature. At Case Western Reserve University, Morton H. Litt developed a polybenzimidazole (PBI) electrolyte for direct methanol fuel cells (DMFC). Researchers found that in addition to being impermeable to methanol, PBI membranes at 80 degrees Celsius have a proton conductivity comparable to that of typical PEMs at the same temperature. However, while a typical PEM's conductivity decreases at over 100 degrees Celsius, the PBI membrane's conductivity continues to increase.

The Solid State Ionics Laboratory under Sossina Haile at Caltech is developing solid acid fuel cells to meet this same goal. Like the PEM electrolyte, the solid acid electrolyte conducts protons but does not need to be hydrated, allowing the fuel cell to operate at temperatures above 100 degrees Celsius. This improves the overall efficiency of the fuel cell while maintaining a relatively quick start-up time and saving the inconveniences of the water management and temperature regulation systems needed by PEM fuel cells. However, the solid acid fuel cell's electrolyte is water soluble and, while experiments have shown that steam is harmless, contact with liquid water could destroy it. Thus, researchers must either design a system to protect the electrolyte from the water during shut down, or develop a water insoluble electrolyte.

TOWARDS CLEAN TRANSPORTATION

Vehicles with fuel cell engines combine the convenient refueling of a combustion engine with the efficiency and cleanliness of rechargeable batteries. Although many obstacles prevent the widespread adoption of fuel cell cars, legislation concerning urban air quality has forced car manufacturers to develop and to market a number of prototype zero emission vehicles (ZEVs). California and some Northeastern states will require that, by 2003, ten percent of all cars sold each year be ZEVs. The combustion of fossil fuels has caused a major increase in the amount of carbon dioxide in the atmosphere, changing the Earth's climate (see Figure 2).

In response to growing concerns about global warming and its effects on the Earth's ecology, in January 2002 the California Assembly approved the first legislation ever to regulate vehicle emission of carbon dioxide. Taking effect in 2005, this legislation is pressuring car companies to devote significant resources towards developing efficient fuel cell vehicles at low cost.

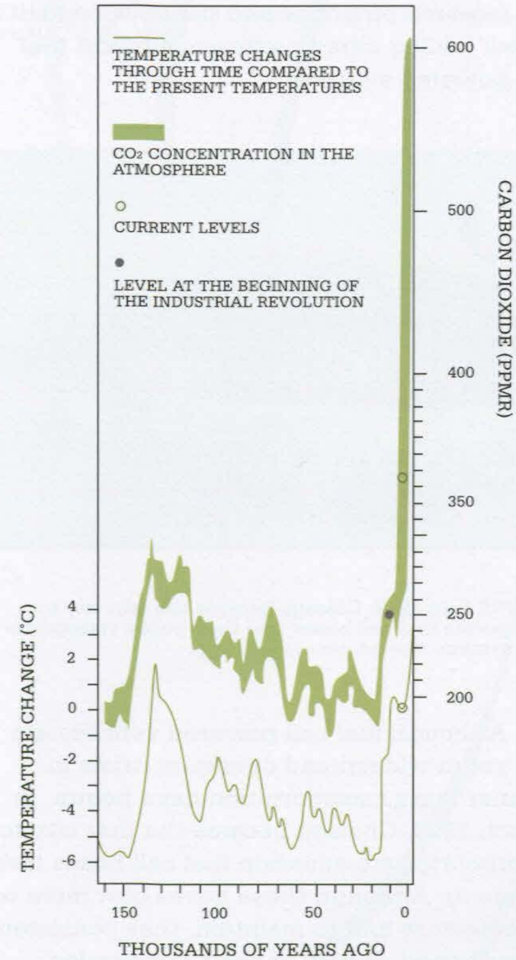


FIGURE 2. The levels of carbon dioxide in the Earth's atmosphere have more than doubled since the Industrial Revolution, causing increased global temperatures. Recent legislation is attempting to slow this process. *Source: Office of Science and Technology Policy*

Many car companies feel that the ZEV mandates present an unfair and heavy short-term financial burden. These mandates will force them to produce unpopular, battery operated vehicles that will soon become obsolete as fuel cell technologies improve. The federal government, eager to end the nation's dependence on foreign oil, has been helping to make this transition period relatively short by supporting fuel cell technology research. The Bush administration has proposed that $150 million of the 2003 federal budget be spent on the FreedomCAR program, which supports the advancement of fuel cell technology by funding research programs and the development of a new fueling infrastructure to support fuel cell powered vehicles.

FIGURE 3. In 1998, Chicago became the first city to incorporate fuel cell buses into their public transportation system. *Source: Ballard Power Systems*

Although fuel cell powered vehicles are not yet in widespread operation, trials in cleaner mass transportation have begun. In March 1998, Chicago became the first city to incorporate zero emission fuel cell buses (see Figure 3). Although these buses cost more to manufacture and to maintain, they consistently performed as well as their combustion engine counterparts without producing air or noise pollution. The California Fuel Cell Partnership, consisting of auto manufacturers and energy companies, intends to have 70 fuel cell powered buses and passenger vehicles operating in California by 2003. The partnership also plans to open four hydrogen refueling stations.

Despite promising initial results, there are still many obstacles to general use of fuel cell technologies. One major question is which fuel to use in future fuel cell vehicles. Although hydrogen is the cleanest and the most efficient fuel, engineers will still have to design ways to efficiently store hydrogen in the vehicle and ways to refuel these tanks. A whole new hydrogen infrastructure would have to be developed, and the public would have to be educated on pressurized gas safety.

Methanol or hydrocarbon fuels such as gasoline may be more convenient choices since the public is already familiar with liquid fuels. However, using these fuels in a fuel cell creates byproducts other than water. Some of these byproducts can degrade the performance of the cell, while others may have harmful effects on the environment. Using other fuels also requires the vehicles to carry onboard units to reform these fuels into hydrogen, adding weight and complexity to the system. Even so, a fuel cell powered engine running on gasoline is more energy efficient than a combustion engine.

Although automakers may not welcome the transition period, the automobile industry continues to improve and to refine fuel cell technology in the hope that breakthroughs will soon lead to a marketable product. Meanwhile, many car companies are taking advantage of this opportunity to redefine many aspects of present day vehicles. In January 2002, General Motors introduced its new concept car, AUTOnomy, overturing preconceived notions of car architecture, which were based on the internal combustion engine, in order to design a car to better suit a fuel cell. The finished product was a 6 inch thick skateboard-like base containing the fuel cells and the power train necessary to run the vehicle (Figure 4). Different car bodies can be

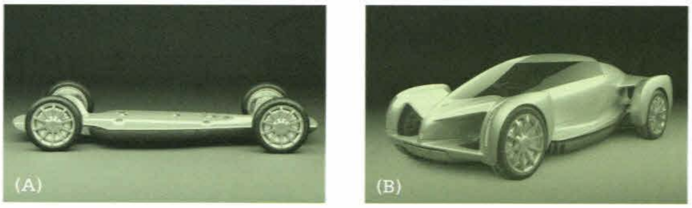


FIGURE 4. (A) The powertrain of General Motor's Concept car, AUTOnomy, fits inside a skateboard-like base. (B) A number of car bodies can be attached to the base. *Source: GM Communications*

attached to the same base, increasing the production efficiency of the fuel cell car while lowering its cost.

In partnership with Delphi Automotive, BMW developed a prototype vehicle that uses an SOFC to run an auxiliary power unit (APU). The APU was used to power the electrical features in the car, reducing the load on the internal combustion engine. As the number of electrically powered devices in a car increases, this innovation can dramatically improve the efficiency of the combustion engine and reduce its emissions. Devices such as air conditioning, radio, and navigation systems powered by the APU can run while the engine is turned off. However, until SOFCs drop in price, fuel cell powered APUs will only be available in luxury vehicles.

SEEKING WIDESPREAD ADOPTION

With the constant improvements in fuel technology, the greatest challenge is to gain public acceptance. The high cost of fuel cell vehicles is likely to be a prohibiting factor. Automobile companies can lower the cost of production by selecting a fuel cell technology to commit to, then building the manufacturing plants to mass produce these fuel cells, but due to the rapid progress in fuel cell technology, car manufacturers are unwilling to invest for fear of rapid obsolescence. However, with the growing financial support of programs such as FreedomCAR and the efforts of scientists and researchers across the world, fuel cells are very likely to become a part of daily life within the next ten years.

Michelle K. Allis is a fourth year undergraduate in Mechanical Engineering at the California Institute of Technology. *The author wishes to thank Dr. S. Haile and G. Pierce for their support and contributions.*

FURTHER READING

1 Los Alamos National Laboratory, *"Fuel Cells: Green Power,"* http://education.lanl.gov/resources/fuelcells/fuelcells.pdf.

2 JT. Norby. The Promise of Protonics. *Nature* **410**, 877-878 (2001).

3 B. C. H. Steele and A. Heinzel. Materials for fuel-cell technologies. *Nature* **414**, 345-352 (2001).

MEASURING DISTANCES BY THE EYEFUL

BY CHRISTINA L. TELLES

IMAGINE COUNTING THE TREES, BUILDINGS and gas stations you pass while driving. Now imagine using those counts to determine how far you have traveled. Although humans use landmarks to help navigate a route, using the frequency of these features to estimate distance seems difficult and imprecise. With-out a keen sense of speed, time, and an accurate count, this estimate becomes a blurry guess. Not so for honeybees.

Feature frequency is precisely the way honeybees estimate the distance they have flown and navigate the terrain they have covered to reach their sources of nectar. Behavioral experiments in the last ten years have elucidated the way in which bees communicate navigation information to one another. Building on Nobel Laureate Karl von Frisch's pioneering 1946 study of honeybee dances, scientists have demonstrated that dances can convey information about the direction or distance to nectar, but the actual mechanism the bees use to estimate distance has only recently been understood.

NEW BEE NAVIGATION THEORIES

Several projects conducted by the research groups of Harald E. Esch at the University of Notre Dame and Mandyam V. Srinivasan at the Australian National University have supported the idea that bees measure how far they have flown by the amount of optic flow they have encountered on the way.

Optic flow is defined as the amount of visual image that has passed through a field of vision–imagine being in a car on the highway and watching road signs and billboards whiz by. Researchers modify the visual input the bees receive in flight by changing the patterns, widths, and motions in the walls of the tunnels which the bees are trained to fly through in order to reach food. The researchers can then observe how those changes affect the bees' perception of how far they have flown. Comparing the actual distance a bee has flown to how far it thinks it has traveled helps researchers understand optic flow. The presence of a visual odometer is a major insight into insect behavior and could lead to comparisons between bees and other more complex animals.

INTERPRETIVE DANCES OF HONEY BEES

Without the capacity for a true language, how do honeybees communicate with one another? After extensively studying honeybees in glass-enclosed vertical hives, Von Frisch and his colleagues claimed that bees told each other where food was located by performing specific dances. The group observed several different dances and correlated each to a different type of information.

They observed a spectrum in the shape of bee dances, which ranged from the round dance to the waggle dance (see Figure 1). He also noticed some variation in the shape of dances, as shown by a comparison between normal and sickled dances.

Based on his extensive study of dances performed by honeybees, Von Frisch also realized that the dances clearly indicated how far the food source was from the hive. The first indicator of distance is the shape of the dance (see Figure 2).

The round dances indicate sources relatively close to the hive, or less than 30 meters away. For long distances, the bees perform more of a "waggle" dance. For intermediate distances the bees perform a more sickled dance. By judging the basic shape of the bee's dance, observers can predict how near or far the food source is located.

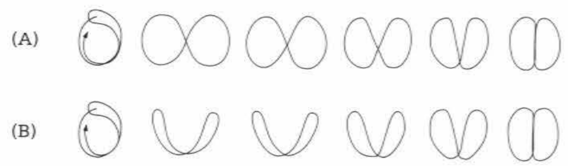


FIGURE 1. Range of Shapes in Bee Dancing, Round to Waggle (left to right): The bee dances follow these stereotypical patterns of movement. The further away the source is from the hive, the more 'waggled' (rightward) the dance becomes. (A) Shows normal dances. (B) Shows sickled ones. *Source: Von Frisch, Karl (1971).*

FIGURE 2. Round vs. Waggle means Near vs. Far: As the distance to the food source increases, the shape of the dance does as well. A rounder dance means closer, a more sickled dance means slightly further away, and the waggle dance means very far away. *Source: Srinivasan, MV, Zhang SW, and Collett, TS. J of Experimental Biology 199. (1996).*

"The honeybee **dances** clearly indicate how far the food source is from the hive"

A more precise way of measuring the distance indicated by a bee's dance is to monitor the duration of the dance. The longer the bee dances in the waggle conformation, the further away the food source is located. Likewise, the more waggles per waggle cycle, (each completed figure counts as one waggle cycle) the further the target is situated (see Figure 4). Unlike the relatively straightforward, widely accepted mechanisms by which the bees communicate the distances they have flown, the methods they use to measure these distances has been a point of contention.

DEBATES OVER HONEYBEE ODOMETRY

Since the advent of the era of the new bee biologists, these revolutionaries have proposed many theories as to how bees measure distance. One of the simplest mechanisms would be for the bee to count something as it flies, perhaps the number of landmarks it has passed or the number of wing beats. However, the simplicity of this mechanism would also allow for a great amount of variation between individual bees, dependent on factors such as the bee's fitness and its particular flying style.

Another proposed mechanism would be for the bee to measure its velocity and the duration of the trip. It could then combine the two to determine how far it has traveled. As it has not been shown that a bee can perform either of these measurements, this theory is unlikely to be true.

The most favored explanation of the early bee biologists was the energy consumption hypothesis, which surmised that the bee could determine how far it had traveled based on the amount of energy it had used. However, a recent resurgence of interest in how bees measure distance has indicated that none of the theories proposed above could be correct.

A group of researchers led by Esch repeated early experiments which had supported the energy consumption hypothesis, and found that the data was less conclusive than first thought. One of the first experiments Esch et al. repeated was Heran's mountain

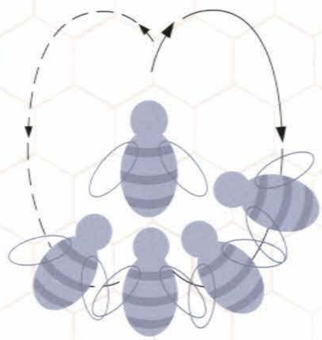


FIGURE 3. The central bee performs his dances while his hivemates gather information mainly through tactile input. *Source: Srinivasan, (1996).*

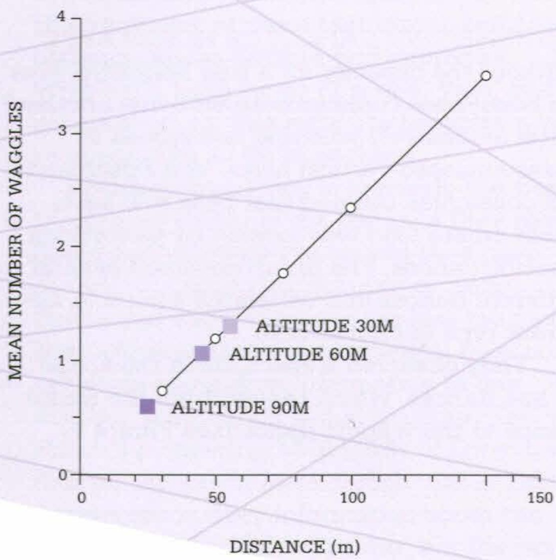


FIGURE 4. Distance vs. Mean Number of Waggles: The number of waggles per dance session has been shown to correspond to the distance to the food source. A longer distance elicits a longer dance. *Source: Esch (1996).*

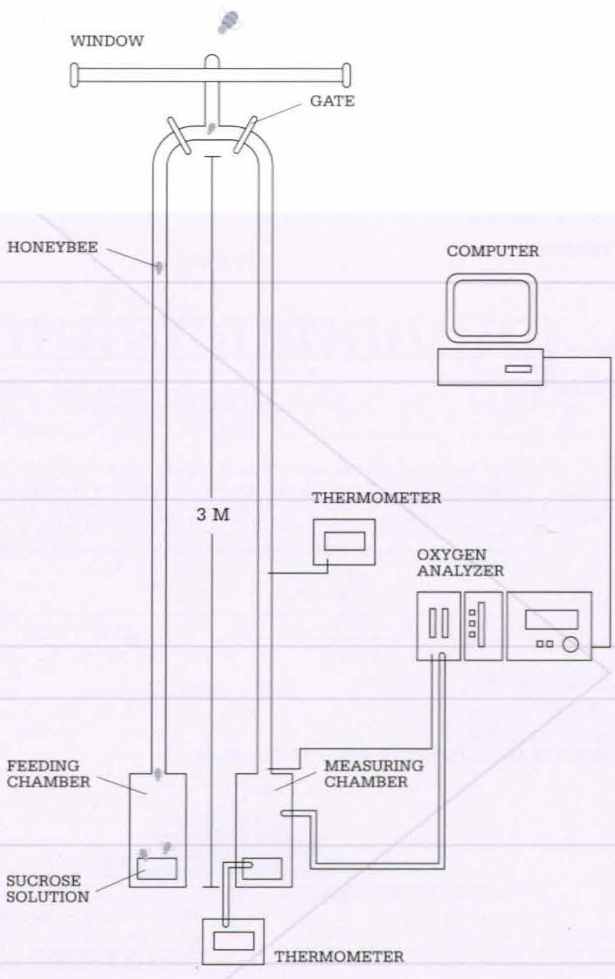


FIGURE 5. Runway Experiment Setup: Bees were trained to walk down a narrow passage, without the option of flying. Energy expenditure was measured via oxygen consumption analysis. *Source: Esch (1996).*

slope experiment. The original experiment involved constructing foraging routes for bees both uphill and downhill of a hive, and then monitoring their dancing to see how far they believed they had gone.

Heran's data indicated that the amount of energy used indicated distance traveled, but when Esch performed a similar experiment he did not obtain comparable results. Despite the large difference in energy consumed between the bees that flew to the high feeder versus those that flew to the low feeder, there were no significant differences in the corresponding dances. The evidence Heran and his colleagues had acquired from their mountain experiment was biased to confirm the heavily favored outcome.

Esch's group also reexamined the Runway Experiment from the earlier bee studies. Since a bee spends more energy walking than flying, a small distance walked might result in a dance indicating a long distance flown.Heran claimed that his data showed that a three meter walk alone resulted in waggle dances indicating flights between fifty and one hundered meters, but when Esch repeated the experiments (see Figure 5), he did not gather similar results. Although he determined (via an oxygen consumption measurement) that the bees had used as much energy in a three meter march as they would in a one hundred twenty-eight meter flight, the bees did not signal that they had covered great distances.

Results were showing fault with the assumption that dance length was related to energy expenditure. Esch designed a new experiment to test the energy hypothesis. He attached a feeder to a balloon, which allowed him to vary the altitude (and thus energy consumed by the bees) while keeping the ground distance to the feeder fixed. Much to his surprise, the dances did not increase in length with the elevation of the balloon but rather got shorter. Energy expenditure had increased, but the perceived distance had decreased. The energy consumption hypothesis was no longer a viable option.

Bee biologists moved to fill the void left by the energy consumption hypothesis with new experimental theories. Visual input appeared to be a crucial factor in allowing the bees to determine the distance they had flown. When scientists trained the bees to fly over an undisturbed lake to a food source, the bees underestimated the distances, if they even made it across the lake. Many of the bees drowned as they moved closer to the water in effort to gain more visual input. The high building experiment lent support to the visual input theory (see Figure 6); the experiment involved locating a hive and a feeder on two separate tall buildings. The scientists then monitored the bees to see if they danced to indicate the actual horizontal distance they had traveled, but the bees flying to the high feeder indicated a much shorter trip. The optic flow hypothesis was gaining support, but could it predict experimentally verifiable results?

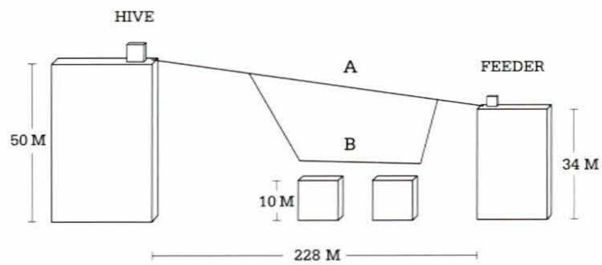


FIGURE 6. The High Building Experiment Setup: Both the hive and the feeder were placed on top of high buildings. The bees were monitored to endure that they took a direct route (A) to the feeder, rather than descending to the ground, flying then ascending to the feeder (path B). *Source: Esch (1996).*

PUTTING THE OPTIC FLOW HYPOTHESIS TO THE TEST

M. V. Srinivasan's group trained bees to fly through a tunnel with a reward at a fixed location (see Figure 7). During the testing periods, the tunnels were modified and then the main search location was determined by taking the mean of the first four positions at which the bee turned around in the tunnel. They used this position to determine where the bee expected to find the reward. The top of the tunnel was covered by mesh, allowing visual cues from the sun to reach the bee while constraining the bee to the tunnel. The first test verified that there was no difference in search patterns between a tunnel lined with a random texture and one lined with stripes running perpendicular to the length of the tunnel (cross stripes). The experimenters used these cross stripes to supply visual stimuli throughout the trials.

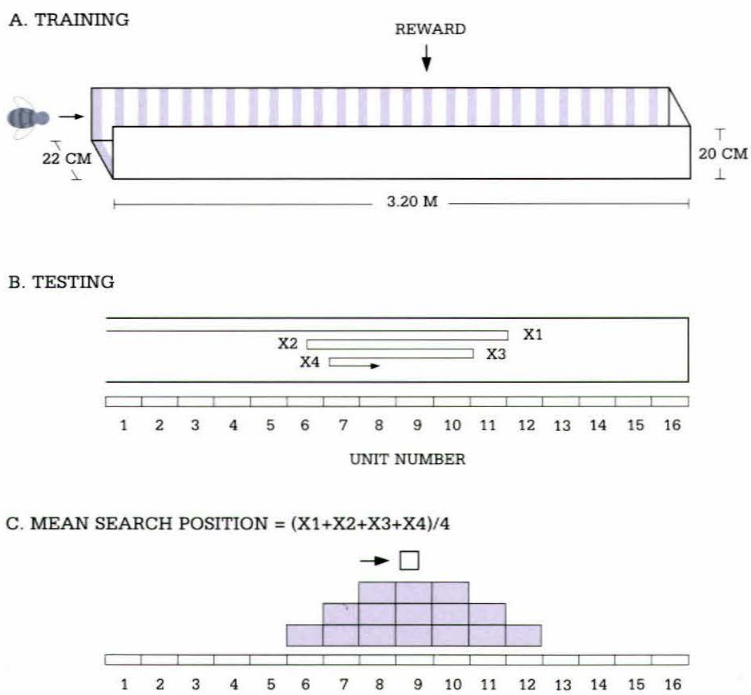


FIGURE 7. Basic Training Protocol: Bees are trained to fly in a somewhat restricted tunnel. The walls and the floor of the tunnel are painted with stripes perpendicular to the path of the bee. The mean search position is then determined by averaging the points where the bee reverses direction. *Source: Srinivasan, MV, Zhang SW, and Bidwell, NJ. J of Experimental Biology 200 (1997).*

The researchers then tested to see if the bees were using the tunnel entrance to gauge how far they had traveled into the tunnel. As one approaches the exit of a tunnel, the aper-

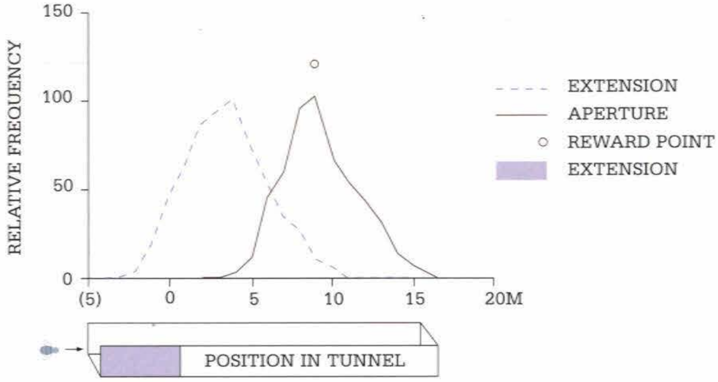


FIGURE 8. Aperture and Tunnel Length Search Patterns: Changes in the aperture does not 'trick' the bees, while the extension does. This indicates the bee knows how far into the tunnel the reward should be, but does not use the reduction in the opening size to judge this distance. *Source: Srinivasan (1997).*

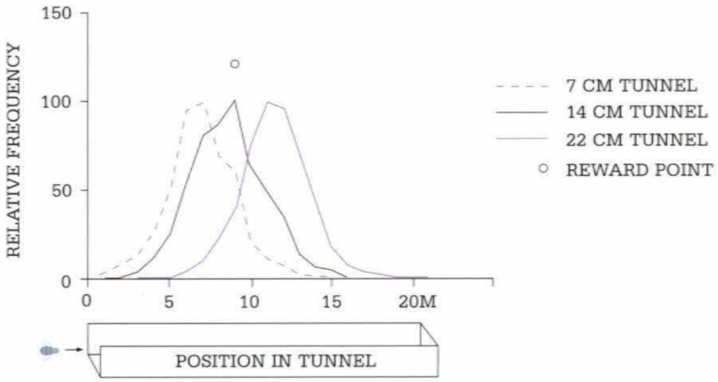


FIGURE 9. Tunnel Width Search Patterns: As predicted by the visual flow hypothesis, the narrower tunnel makes the bees think that the reward is closer to the end of the tunnel, or that they have traveled further. Likewise, they search deeper into the wider tunnel for the reward. Source: *Srinivasan (1996, 1997).*

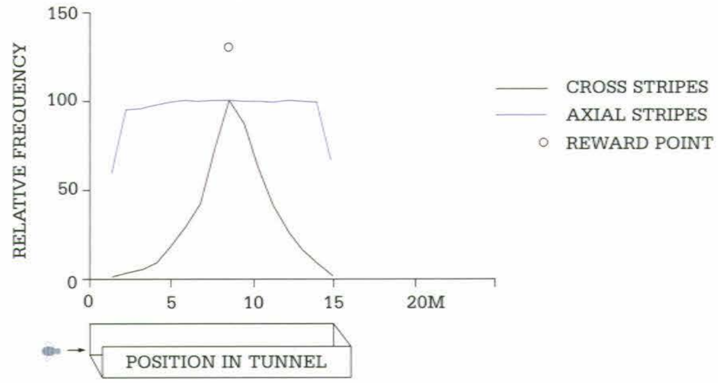


FIGURE 10. Stripe Orientation and Search Pattern: Changing the pattern that covers the inside of the tunnel to axial stripes (along bee's path), the bee searches equally along the length of the tunnel. Axial stripes offer very little visual flow input, so this result supports the hypothesis. *Source: Srinivasan (1997).*

ture appears to grow; likewise, as one moves away from one end, the aperture appears to shrink. To see if this effect was guiding the bees' perception of distance, Srinivasan and his colleagues modified the apertures of the tunnel ends, but found that doing so caused no significant change in the search patterns. However, when they extended the front end of the tunnel, the bees altered their search patterns to compensate for the extra length they covered (see Figure 8). They concluded that the bee used the entrance of the tunnel as a landmark to measure the distance to the reward. These results are consistent with the optic flow hypothesis.

The next experiment tested the effect of tunnel width on the search pattern of a trained bee. When you wave your hand back and forth close to your face, the amount of movement seems much more significant than it does when waving at arm's length–the optic flow is much greater when the stimulus is closer. The optic flow hypothesis predicts that the bee should search at shorter distances when the tunnel is narrower and at longer distances when it is wider. The experiments verified these predictions (see Figure 9).

The most striking experiment involved changing the stripes on the inside of the tunnel from cross stripes to axial stripes (parallel to the length of the tunnel). Just as the lane markers on the freeway blur as you travel parallel to their path, the bees could not detect much difference between the axial stripes as they flew by. Figure 10 shows that bees in the axial-striped tunnels searched along the entire length almost equally, indicating they had no idea how far into the tunnel they were.

Srinivasan and his colleagues then conducted a more quantitative study. Four different experimental setups were designed, as seen in Figure 11. Tunnels 1, 2, and 3 were set up with their entrances 35 meters from the hive, while tunnel 4 was positioned 6 meters from the hive. Tunnel 1 had the reward at the near end of the tunnel, while the other three had the reward at the far end of the tunnel

(6 meters inside). Tunnel 3 had axial stripes along its length, while the other three tunnels had cross stripes.

Considering the distances from the hive to the tunnels, one might expect that the bees would signal the distances to the rewards with round dances. However, the two tunnels with the reward at the far end and with the cross stripes elicited largely waggle dances, indicating estimated distances of at least 50 meters. This is linked to the bees' receiving a greater amount of optic flow while traveling through those tunnels. Likewise, in the tunnel with the reward near the entrance and in the axially-striped tunnel, the optic flow was not deceptive and the dance reflected the accurate distance.

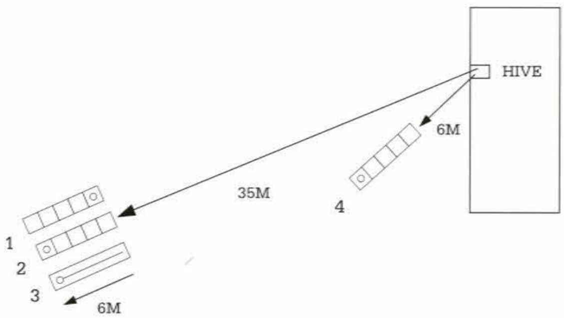


FIGURE 11. Visual Odometry Setup: Placement of various tunnels. Tunnels 1-3 are at 35 meters from the hive, while 4 is at 6 meters. The reward is at the back of the tunnel for all tunnels except 1. The pattern in the tunnel is cross-stripes except tunnel 3. *Source: Srinivasan (2000).*

CONSEQUENCES OF VISUAL ODOMETRY

Visual odometry not only accounts for foraging behavior, but also flight behavior. It has been shown that bees fly at a relatively constant speed under normal, open conditions, but decrease their speed when passing through narrow passages. Visually-mediated odometry would account for this behavior. It would also explain why a bee approaching a landing site slows down. As the bee gets closer to the landing site, the optic flow increases in speed due to the decreased distance–slowing down the velocity would then compensate for the perceived increase in optic flow and hold it constant. Such a system would account for smooth landings on small areas.

It is also possible that other insects and animals measure how far they have traveled using visual odometry. Cataglyphis ants rely on moving floor pattern effects to judge distance. In these ants, the speed of image flow is more important than frequency, as patterns of different temporal spacing did not affect distance perception.

Visually-mediated odometry, or the optic flow hypothesis, explains the results of these experiments. Experimental confirmation that some species use this method of navigation helps researchers to make hypotheses about how other animals might navigate their surroundings and raises more questions such as how these animals choose and keep track of the features that pass by the eyeful.

Christina L. Telles is a fourth year undergraduate in Biology and Literature at the California Institute of Technology. *The author wishes to thank Professor Masakazu (Mark) Konishi and Carol Readhead.*

FURTHER READING

1 H. E. Esch, S. Zhang, M. V. Srinivasan, and J. Tautz. Honeybee dances communicate distances by measured optic flow. *Nature* **411**, 581-3 (2001).

2 Collett, T. Measuring Beelines to Food. *Science* **287**, 817-818 (2000).

3 M. V. Srinivasan, et al. Honeybee Navigation: Nature and Calibration of the Odometer. *Science* **287**: 851-853 (2000).

PRIVACY THROUGH UNCERTAINTY: THE SECURITY OF QUANTUM ENCRYPTION

BY DUNCAN JOHNSTON

QUANTUM CRYPTOGRAPHY PROVIDES A NEW SYSTEM FOR securely transmitting a message from one location to another. Current popular cryptography methods are based on modular arithmetic algorithms, such as the one developed by Ron Rivest, Adi Shamir, and Len Adleman in 1978. The electronic key used in these algorithms can be decoded and the security can therefore be compromised by exhaustive computer searches. Using current computers, this search would take an unfeasibly large amount of time, but a working quantum computer could break the key within a few minutes. Although a practical quantum computer has not yet been constructed, many physicists are working toward that goal. These computers would be capable of truly parallel computations and could use algorithms developed by Peter Shor and Lov Grover to break a key much faster than a conventional computer. This possibility presents the need for a new and completely different approach to encryption, which is where quantum cryptography comes in.

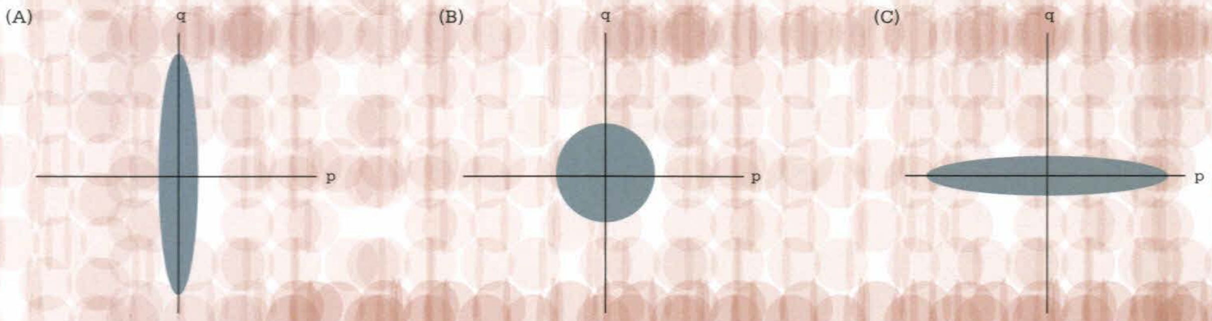


FIGURE 1. Different ways of squeezing: (A) Squeezing the photon state in p (B) No squeezing. (C) Squeezing in q.

We have studied the security of one particular quantum cryptography system–quantum key distribution using squeezed states. This system was originally proposed by Mark Hillery in 2000, and was further developed by Daniel Gottesman and John Preskill. The system studied here is a simplification of that scheme, allowing us to make a quantitative analysis of the parameters required for security.

The security analysis is carried out by imagining an eavesdropper who has the ability to make two different types of measurements on a signal. It is important that the quantum cryptography system can be tested to ensure that it is secure against both types of eavesdropping.

ESTABLISHING A SECURE KEY

The basis of quantum cryptography is the technique of quantum key distribution (QKD). QKD is a method by which an electronic key is created between two users (usually called Alice and Bob respectively). This key can then be used to encode secret information before it is sent from Alice to Bob over a communication channel that could potentially be eavesdropped on by another person (called Eve).

The type of key most normally used in QKD is a simple binary addition key. For example, consider an 8-bit message, "10010110," which Alice wishes to send to Bob. Alice and Bob share an electronic key of the same length as the message. Let us say that this key is "10101010." Alice performs the bitwise addition (where 1 + 1 = 0, 0 + 0 = 0, 1 + 0 = 1, and 0 + 1 = 1, also known as an XOR operation) of the key and the message. This produces the encoded message "00111100." When Bob receives the encoded message, he performs an XOR operation between it and the key to retrieve the original secret message, "10010110." This system is called a one-time pad because the key can only be used once and must then be discarded.

Provided that Alice and Bob never re-use the same key, they can be sure that Eve cannot gain any information about the secret message simply by eavesdropping on the encoded messages, which could be sent over a public channel such as the Internet. The challenge of quantum cryptography is to provide a way for Alice and Bob to create a key while being certain that Eve has no information about that key that will be useful to her.

All QKD protocols involve two non-commutable, or orthogonal (different), bases in which a single bit of information can be encoded. If a measurement of the signal is made in the wrong base, the result is probabilistic according to the laws of quantum mechanics and gives no useful information.

Squeezed state QKD encodes information by squeezing a coherent state in one of two orthogonal phase quadratures, p or q (see Figure 1). A squeezed state is a photon state in which the uncertainty in one phase quadra-

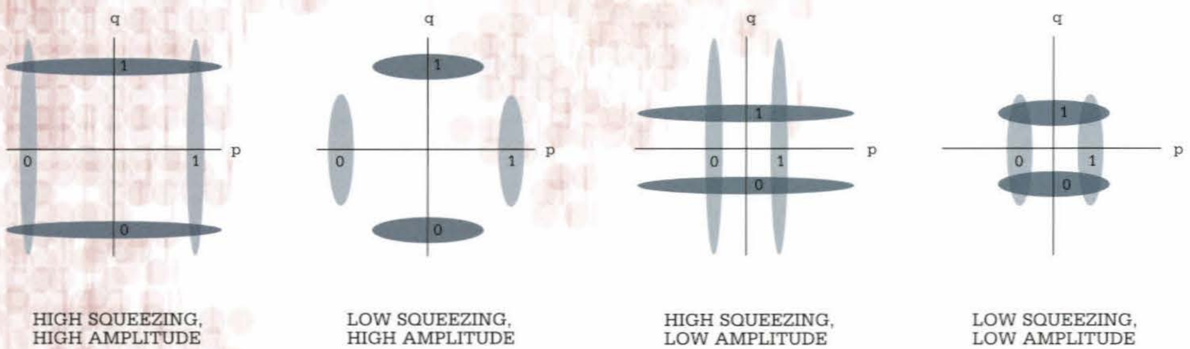


FIGURE 2. Four plots in phase space of the four different squeezed states which are used to encode a bit. Each plot shows a different degree of squeezing or amplitude. The bit has a value of 0 or 1 and is encoded by squeezing in one of the phase quadratures, p or q.

ture (representing a phase component in or out of phase with a fixed reference) has been reduced at the expense of increasing the uncertainty in the orthogonal phase quadrature. The value of the bit being sent is then encoded by translating the state along the axis in which it has been squeezed. The security of the system relies on the fact that states squeezed in p will have a significant overlap with states squeezed in q (see Figure 2). The squeezed state can be described by its amplitude, a measure of the strength of the signal, and by the squeezing parameter, a measure of how much the signal is squeezed.

Bob chooses to make a measurement of either p or q at random. If he makes his measurement along a different quadrature than the one in which Alice had squeezed the state, he will not be able to gain any information that will be useful to him.

Alice and Bob both randomly choose a phase quadrature to use for each bit. After all of the quantum information has been transmitted, they can use a public channel to compare which quadratures they used for each bit and then discard information for which they used different quadratures. The remaining information is retained to produce the cryptographic key (see Figure 3).

If Eve attempts to make a measurement of the message and then re-transmit it, she will not know whether to measure the value of p

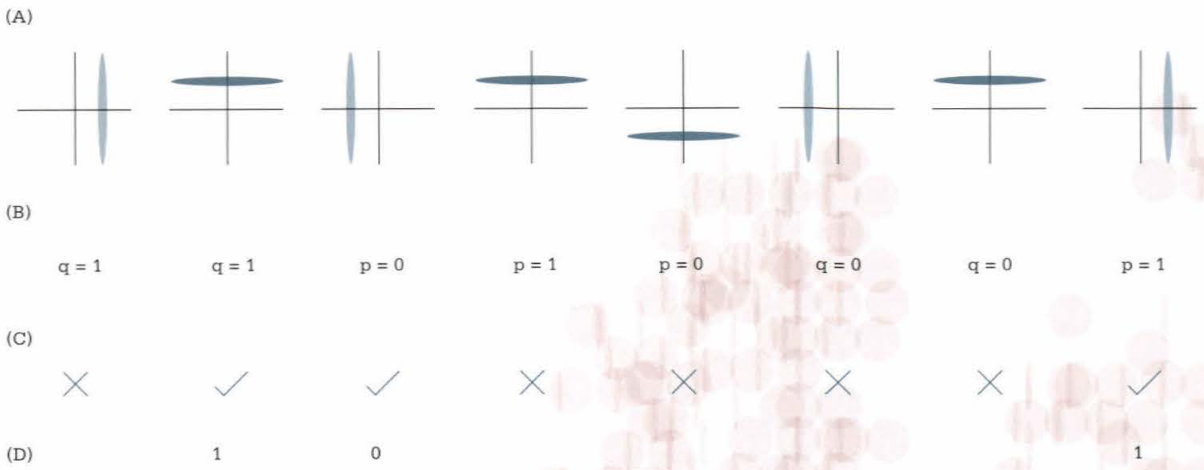


FIGURE 3. Generating the cryptographic key. (A) Alice generates a sequence of bits, randomly choosing between squeezing in p and q. (B) Bob randomly chooses to measure in p or q, and notes the value he measures. (C) Alice and Bob use a public channel to compare the bases they used for each bit and keep only the bits in which Bob read the correct base. (D) This gives a final key of 101.

or q and will therefore introduce an error rate in the signal, which Alice and Bob can detect by comparing a part of their keys later.

"If Eve uses a **homodyne** detection technique to eavesdrop, she has to make a random decision as to whether to measure p or q..."

MEASURING THE SIGNAL

The system outlined above involves Bob's making a homodyne measurement along one phase quadrature in order to determine the amplitude of the signal. This involves interfering the source signal with a local oscillator of the same frequency as the signal and measuring the amplitude of the interference pattern. It is necessary to know the phase of the signal in advance for a homodyne measurement to give any useful information. This is suitable for Bob, since he assumes that the signal is squeezed in one of two possible phase quadratures and only considers the results from measurements in which he chose the same quadrature Alice did.

If Eve uses a homodyne detection technique to eavesdrop, she has to make a random decision as to whether to measure in p or q, after which she re-transmits the bit according to what she detected. This method introduces a fixed error rate in the Alice–Bob channel irrespective of the overlap of the states used by Alice and Bob (see Figure 4).

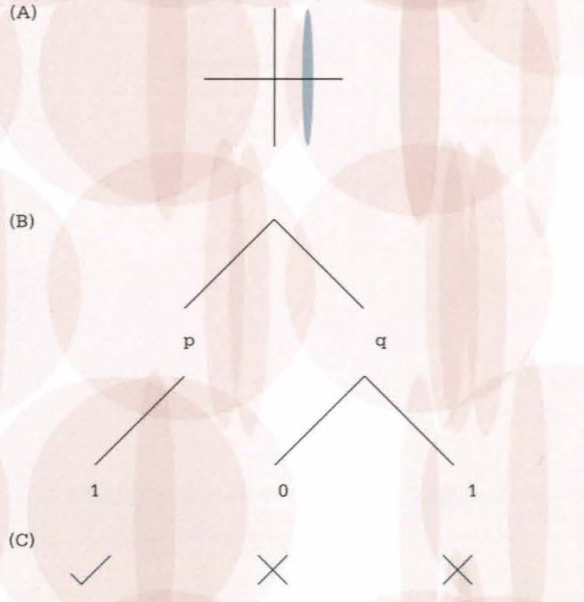


FIGURE 4A. When Alice transmits a bit to Bob without someone eavesdropping, the error rate is the one intrinsic to the communications system. (A) Alice sends a bit (in this case, p=1). (B) Bob could measure the bit in p or q. If he measures in p, it is likely that he will read 1, but if he reads in q, he could read 0 or 1. (C) Alice and Bob compare the bases they used. If the bases match, they keep the bit. If they don't match, they discard the bit. The error rate in this case is only the intrinsic error rate of the system.

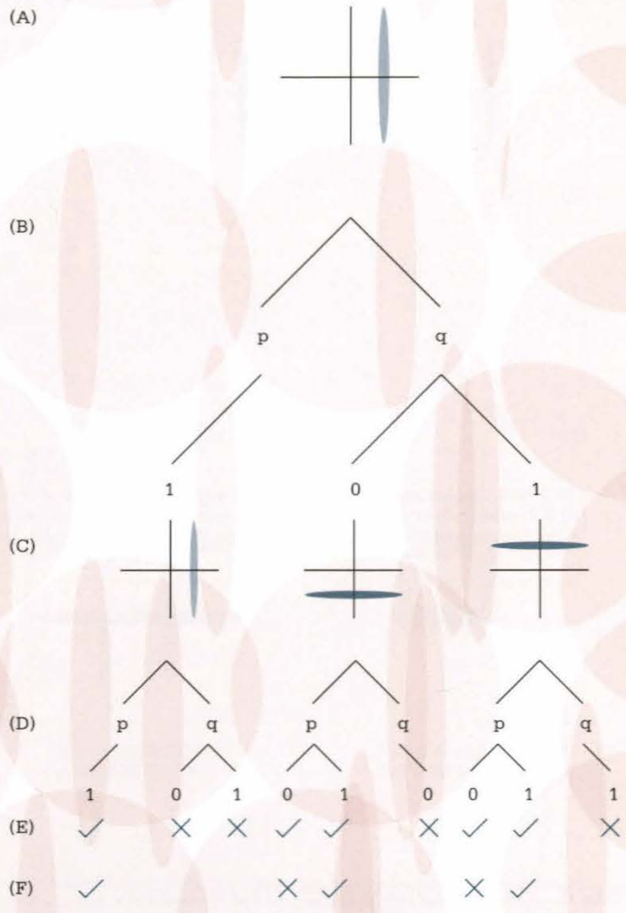


FIGURE 4B. When Alice transmits a bit to Bob with an eavesdropper, the error rate increases. (A) Alice encodes a bit (again, p=1). (B) Eve intercepts the signal and makes a homoodyne measurement in p or q. If she reads in p, she will most likely read 1, but if she reads in q, she could read 0 or 1. (C) Eve retransmits the bit she read to Bob. (D) Bob could read the bit in p or q. (E) If Bob reads in p after Eve retransmitted in q, Bob and Alice could agree on the base, while disagreeing on the actual value of the bit. This additional 25% error rate was introduced by Eve.

An adaptive phase measurement such as the one devised by Howard Wiseman and Rowan Killip can make an efficient measurement of phase with no prior knowledge of the signal. The result of this measurement gives an estimate of the phase of a signal but does not give any information about the amplitude. If Eve could make such a measurement, she could gain some information about the key without being detected by Alice and Bob. It is assumed here that Eve is able to make an adaptive measurement of phase according to the Wiseman technique. This would enable her to distinguish, within the limits of the quantum uncertainty, between the four different possible signal states.

Eve could choose to eavesdrop on only a certain fraction of the bits transmitted by Alice. She would not be able to obtain information about the entire key, but would introduce a smaller disturbance in the Alice–Bob channel. This could enable Eve to gain some information about the key without Alice and Bob noticing, in which case they would think the cryptographic key they had created was safe to use.

The challenge for Alice and Bob is to operate the system such that they can eliminate any possibility of Eve gaining any information about the key. We will see that if Alice and

Bob know that Eve could only have a limited amount of information about the raw key, they can reduce this information to zero by using quantum privacy amplification techniques.

HOW SECURE IS OUR KEY?

Alice and Bob need to have a method of quantifying the security of a channel and how much information can be transmitted along it. They can measure the error rate of the communication channel by publicly comparing a small amount of the key they produce and then use that error rate to deduce whether the channel is secure.

The secrecy of the cryptographic channel is analyzed by plotting the relation of the information capacity to the error rate measured by Alice and Bob. The measured error rate could be affected by two separate factors. Eavesdropping of any degree causes a disturbance and increases the error rate. The states used for transmitting information will also have a finite overlap which will cause an intrinsic error rate. We can calculate the security of the squeezed state QKD system at a given signal amplitude and squeezing parameter by comparing the actual error rates to those expected with and without eavesdropping.

The information capacity is a measure of the amount of information that two parties can share over a given communication channel, as defined by Shannon's coding theorem. It shows how many bits of information Alice could transmit to Bob on average per pulse along a channel with a known amount of noise. A perfect binary channel has an information capacity of one bit per pulse, but when noise is introduced into the system, the information capacity of this imperfect channel is reduced to less than a bit per pulse.

In studying this system, there are two significant information capacities, the information capacity of the channel from Alice to Bob (IAB) and of the channel from Alice to Eve (IAE). Both channels are binary symmetric channels, so their information capacities are as given above. If IAE is less than IAB, then Eve has less information about what Alice sent than Bob does. Using privacy amplification algorithms, it is possible for Alice and Bob to produce a final key from the original key that can be guaranteed as secure. The secrecy capacity, the amount of secure information they can extract from this partially insecure channel, is given by the difference between IAB and IAE. It is possible to calculate the secrecy capacity of a channel for given signal amplitudes and degrees of squeezing for both cases of eavesdropping described above.

The states shown in the figures depict the first standard deviation of the squeezed state, representing the result that would be obtained by a completely efficient measurement in 68% of cases. Areas of lower probability exist outside this area and some will extend beyond the origin of the phase space diagram. This is referred to as the overlap between the 0 and 1 states and it is greater when the amplitude of the encoding states is lower. This means there is a finite probability of Bob measuring a signal as a 0 when Alice transmitted a 1 and vice versa, producing an intrinsic error rate. The remainder of the error rate is due to disturbances caused by Eve's measurements. The total error rate is the sum of these two error rates. The information capacity between Alice and Bob, IAB, is then calculated from this measured error rate. The probability of Eve making a correct measurement of the signal can be calculated in relation to the error rate for eavesdropping using homodyne detection and using adaptive phase measurement, this can then be used to calculate the information capacity IAE.

Alice and Bob need to be sure that they would notice if Eve had eavesdropped on the transmission of a key. Alice and Bob can directly measure the error rate in their channel by comparing a few bits of the cryptographic key on a public, yet reliable (low error rate), channel. This could be any classical telecommunications channel, such as the Internet. They must be able to determine whether the key is secure from this information alone.

"The least secure region is that where the amplitude is large, but there is very little squeezing."

With either eavesdropping strategy, Eve will introduce a larger error rate if she eavesdrops on a greater proportion of the transmitted signal. Alice and Bob can compare the secrecy capacity they expect at a certain error rate if all of the observed errors are caused by Eve with the information capacity at that error rate if the observed errors are all random errors. This ratio of secrecy capacity to the information capacity of a secure channel with the same error rate is called the relative secrecy. This ratio gives a quantitative measure of the channel's efficiency in conveying secret information.

Alice and Bob would prefer a large relative secrecy for their channel, as this would enable them to maximize the amount of information that can be transmitted. They also need to be able to determine if the relative secrecy has decreased to the point that the channel is no longer secure against attacks from Eve. This occurs when the measured error rate is close to the error rate at which the relative secrecy would be zero. Due to random errors, there is constraint on how closely Alice and Bob can reliably distinguish between a given error rate and this error rate limit.

The measure of the guaranteed security of a channel is the gradient of the graph of relative secrecy against measured error rate in the Alice–Bob channel. A large gradient indicates a channel with a low security, whereas a smaller gradient indicates a channel with greater security.

OPTIMIZING THE SECURITY

We calculated the gradient of the graph of relative secrecy against measured error rate at the point where the relative secrecy was zero for a range of squeezing parameters and squeezed state amplitudes. In Figure 5(A), we see the log of the gradient where Eve uses the adaptive phasemeasurement eaves dropping technique. Figure 5(B) shows the result where Eve eavesdrops using a homodyne detection technique, randomly choosing between measuring in p or q.

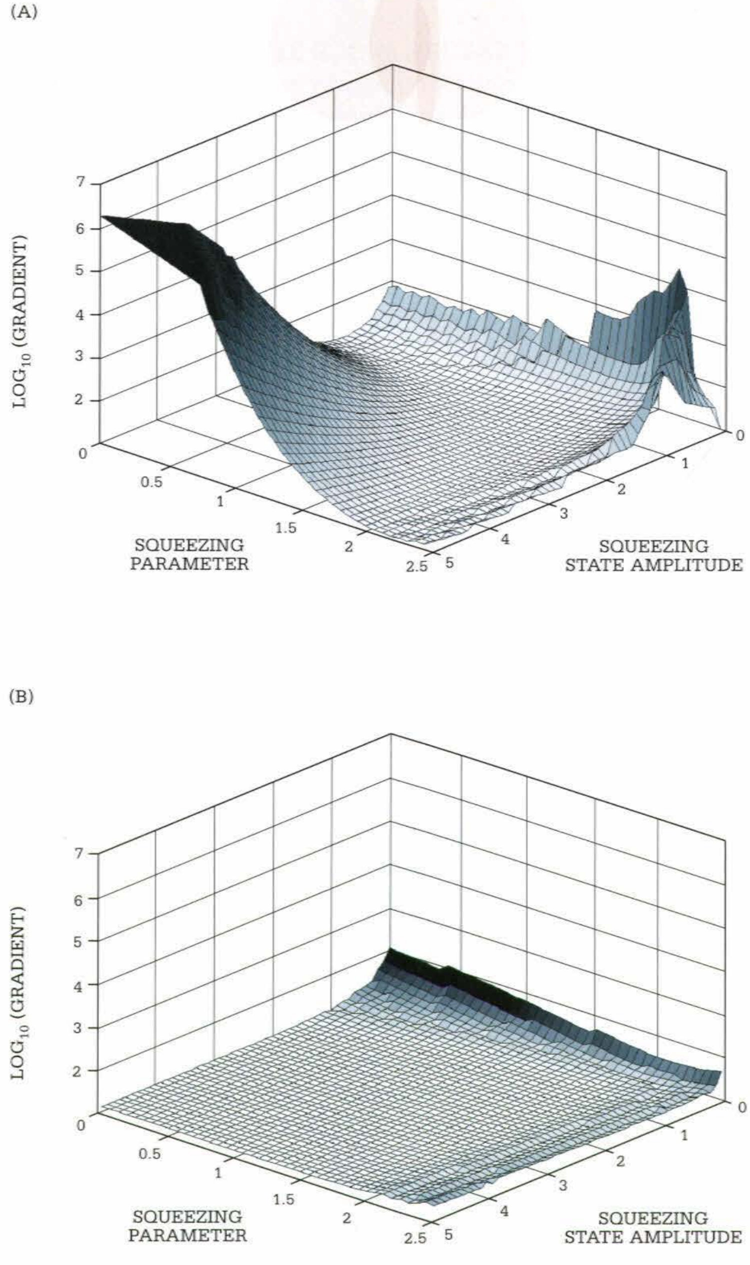


FIGURE 5. Graphs of the gradient of relative security against error rate over a range of squeezed state amplitudes and squeezing parameters when there is an eavesdropper who is (A) using the adaptive eavesdropping technique or (B) using the homodyne eavesdropping technique.

It is easier for Alice and Bob to establish a secure key in the areas where the gradient is low. The safe range of error rates in which they can most easily establish a secure key is reasonably well separated from the point at which the relative secrecy goes to zero. If Alice and Bob communicate in a situation where the relative secrecy is near zero, they will need to sacrifice a large portion of the transmitted information in order to produce a secure key. This would reduce the key creation rate, and therefore slow secure communication.

The two graphs clearly show that the squeezed state QKD protocol is more vulnerable to attacks when Eve is using adaptive phase measurement than to attacks using homodyne detection, since the gradient in Figure 5(A) is greater by up to six orders of magnitude than that in the graph showing security against homodyne eavesdropping.

Figure 5 shows that the least secure region is that where the amplitude is large, but there is very little squeezing. As seen in Figure 2, the states squeezed in p do not overlap significantly with the states squeezed in q. Therefore, it is possible for Eve to differentiate accurately between the four states using an adaptive measurement and then to retransmit the signal, introducing very little disturbance into the Alice–Bob channel.

The most secure area runs from points of low amplitude and low squeezing parameter up to points with amplitude of five and squeezing parameter of two. The region with highest amplitude and highest squeezing parameter would be a useful point at which to operate the system, as signals of high amplitude can be transmitted over a greater distance than signals of low amplitude. Current QKD systems are limited to using single photon states and can only obtain ranges of up to 48 km. By using larger amplitudes, squeezed state QKD could potentially be used between continents. Squeezed state quantum key distribution shows potential as a viable system, but it has not yet been implemented experimentally. Until experiments are conducted, it is difficult to say which parts of the system will be complicated to implement and whether squeezed state quantum distribution will give significant increases in reliability over other techniques. Quantum cryptography is still far from joining desktop computer technology, but it may be seen in military or large commercial applications within the next few decades.

Duncan Johnston is an undergraduate at Corpus Christi College, Cambridge University in England. *This work was completed with Hideo Mabuchi, Associate Professor of Physics at the California Institute of Technology, and funded by the 2001 Caltech Summer Undergraduate Research Fellowship. The author wishes to thank Hideo Mabuchi, Lauren Stolper, and the Cambridge – Caltech exchange program.*

FURTHER READING

1 R. L. Rivest, A. Shamir, and L. M. Adleman. A Method for Obtaining Digital Signatures and Public Key Cryptosystems, *Communications of the ACM* **21**, 120 (1978).

2 M. Hillery. Quantum cryptography with squeezed states. *Phys. Rev.* 022309 (2000).

3 D. Gottesman and J. Preskill, Secure quantum key distribution using squeezed states, http://www.lanl.gov, quant-ph/0008046 (2000).

4 H. M. Wiseman and R. B. Killip. Adaptive single-shot phase measure ments: The full quantum theory. *Phys. Rev.* **A 57**, 2169-2185 (1998).

5 A. K. Ekert, B. Huttner, G. M. Palma, and A. Peres. Eavesdropping on quantum-cryptographical systems. *Phys. Rev.* **A 50**, 1047-1056 (1994).

MUTATION AND COMPETITION:

WHY STICKING TO YOUR ANCESTORS IS A GOOD IDEA

BY JIALAN WANG

THE SCOTTISH GENETICIST J. B. S. HALDANE once remarked that he would gladly drown if by doing so he would save two siblings or eight cousins. Each of his siblings would share one half his alleles; his cousins, one eighth. They could potentially add as many of his alleles to the gene pool as he could. This argument would be strikingly logical if the number of alleles passed on were the only criteria for genetic success. This is analogous to the argument that replication rate—how many offspring a single organism could produce per unit time—is the only factor in genetic success. A new theory of fitness might show that this is not the case.

FITNESS, A NEW DEFINITION

The precise definition of genetic fitness is elusive. It is often broken into parts such as direct and indirect, number of alleles passed on, and number of alleles identical to its own that are passed on. We investigate this complex notion by studying the simplest possible systems: populations of asexual organisms. In this system, the notion of fitness reduces to "the organism's ability to propagate its genetic material," which, in all Earthly biochemical life, comes in the form of DNA or RNA.

This definition of fitness yields two subtle but critical implications. The first is an organism's fitness cannot be measured instantaneously or by plucking it out of a population and performing measurements on it. To determine the ability of a given organism to propagate its genetic material, we must examine the long-term success of the lineage founded by that given organism. The success of the lineage can only be determined after many generations of reproduction, long after the original organism has died.

The second implication is that fitness depends on the environment. Every organism has environments in which it will flourish and environments in which it will perish. An example of this is the diversity of the finches of the Galapagos, first described by Charles Darwin. Each species of finch is adapted to a very specific niche and environment. Although they survive well in their native habitats, they might be poorly equipped to deal with other environments. The finches have beaks of varying shapes depending upon where they live. Ground finches have crushing beaks that are effective in consuming seeds as opposed to tree finches, which have sharper beaks that are better suited to grasping insects.

Evolutionary biology has concentrated on replication rate; however, evolutionary biologists have recently acknowledged a new factor of fitness, robustness. The robustness of an organism is a measure of how quickly the replication rate decreases as a function of distance from the organism's peak fitness. A careful comparison of the importance of replication rate versus that of robustness answers the question of evolutionary success.

QUASISPECIES MODEL

Asexual organisms, whose simplicity offers a unique vantage on evolutionary processes, reproduce by splitting their genetic material into two offspring. Even with this reproduction method, the genes of the offspring are not identical to those of the progenitor. An organism's genetic material usually has a peak in replication rate centered around a group of very similar organisms, the quasispecies.

In 1971, Manfred Eigen developed the quasispecies model; if the ancestor is robust, that is, able to pass on its genes with few mutations, then its descendants are likely to be in its quasispecies. The genetic fitness of the organism depends on the long term success of its descendents. Therefore, it must produce not only enough offspring, but offspring that will be able to pass on genes within the cluster of similar organisms. It follows that an organism must reproduce without its genes mutating more than a set amount in order to produce offspring who carry its genes. This model is based on the fact that only a small fraction of all possible organisms in genotype space are viable (able to reproduce) and points to the importance of robustness for genetic fitness. Figure 1 depicts the fitness peaks of two quasispecies.

The quasispecies theory dictates that when two individual organisms of varying robustness are put in an environment with a low mutation rate, the one with the higher replication rate will produce more descendents. However, if these two organisms are placed in an environment with mutagenic factors such as high-energy radiation or reactive chemicals, the more robust organism will be able to reproduce with minimal mutations to within the peak genetic replication zone. This allows the more robust organism's descendents to multiply more rapidly.

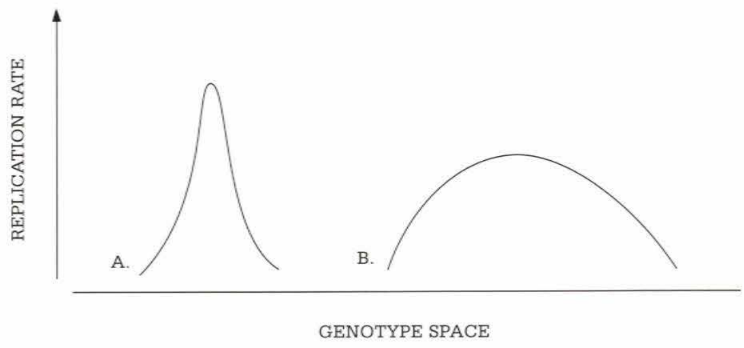


FIGURE 1. Fitness peaks of two quasispecies. Here, quasispecies A has a higher peak than quasispecies B, but has a sharper peak. This means that while the peak genotypes of quasispecies A replicate faster than the peak genotypes of peak B, quasispecies A is less robust.

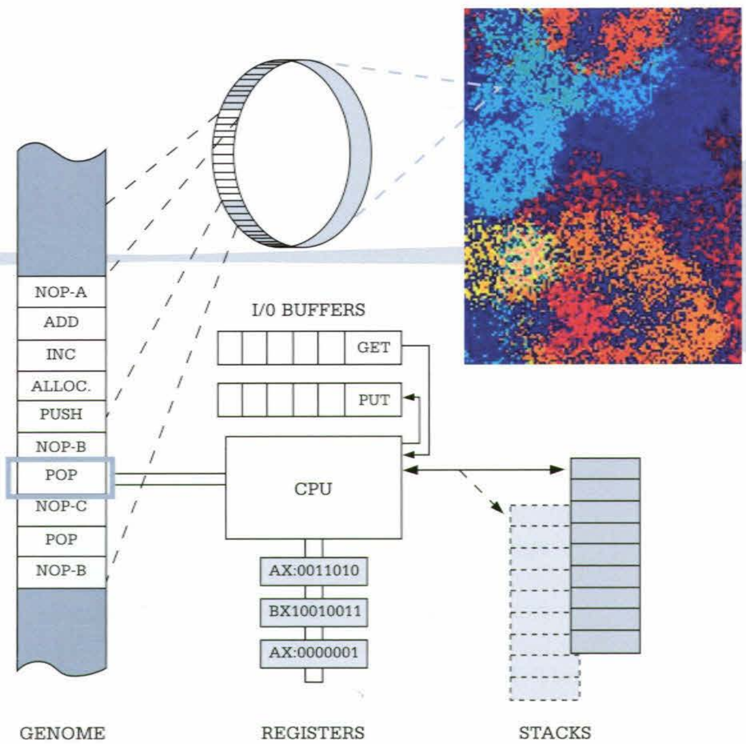


FIGURE 2. A schematic diagram of a digital organism in a population. The colored grid on the upper right represents a population, with the different colors identifying different genotypes. An organism consists of a series of simple instructions, such as those displayed above the heading "genome." Each organism has its own virtual processor, three registers (places to store data), and stacks (places to store instructions) with which it carries out its metabolism of logical operations.

"Digital organisms open many new doors for genetic research."

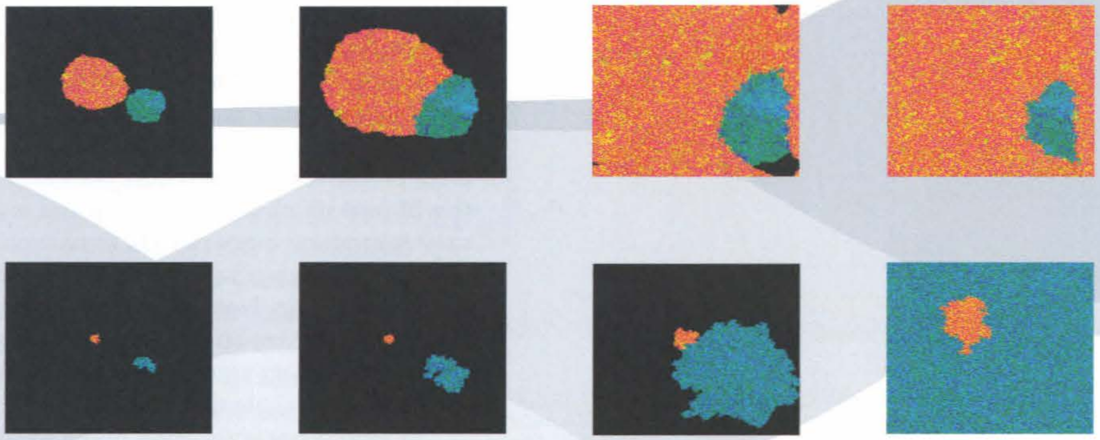
DIGITAL ORGANISMS

Comparing the genetic fitness of two individual organisms is a complicated matter. Fitness itself relies on many factors including size, ability to gather food, and capacity to escape from predators. The environment also affects the fitness of an organism. A given organism may reproduce more readily in certain habitats, which may give an unfair advantage to one test subject. How then can experiments show whether replication rate is the only factor affecting long term genetic success or if robustness also plays a role?

Digital organisms open many new doors for genetic research. Using computer generated organisms researchers can make generalizations beyond the organic forms that have been studied to date. While experiments involving *E. coli* allow a researcher to study hundreds of different genotypes, computer simulations allow researchers to study billions of genotypes at a higher precision, because computer models eliminate certain experimental errors and inconsistencies. The growing complexity of mathematical problems creates interest in the use of digital organisms.

Avida is a platform used for this type of computer simulation. This software generates and maintains a constant population of 3600 organisms. Each organism's genetic code is a sequence of around twenty instructions. Each set of twenty instructions draws from a possible twenty-eight, which roughly correlate with the twenty amino acids. The digital organisms replicate with a pre-programmed probability of mutation, in which case a random instruction is substituted for the mutated one. This models how DNA base pairs can be misread in the event of a mutation.

The simulation allocates resources to the organisms based on population size. Behaviors that increase fitness, such as gathering food and escaping danger, are modeled by giving the organisms tasks to perform. Organisms that successfully complete these tasks receive additional resources which allow them to attempt more tasks and increase their population.



(A)

(B)

FIGURE 3. (A) and (B) show snapshots of two competition experiments between the same two starting organisms at different mutation rates, with the mutation rate in (A) being one third that of (B). Members of the lineage with higher robustness are colored blue, and members of the lineage with higher replication rate are colored orange. The four snapshots of the two growing populations during a competition experiment illustrate the dominance of the lineage with higher replication rate at the lower mutation rate and the dominance of the lineage with higher robustness at the higher mutation rate, as predicted by the quasispecies model.

ONE ANCESTOR AGAINST THE OTHER

This experiment made use of two organisms entitled ancestor A and ancestor B. The organism with higher replication rate and lower robustness, ancestor A, competed against organism B, the organism with a lower replication rate and higher robustness.

The traditional model predicts an overwhelming dominance of ancestor A, because it posits that replication rate is the only relevant factor in this experiment. The quasispecies model, however, predicts that ancestor A will only dominate at relatively low overall mutation rates. At higher mutation rates, ancestor B should dominate the population. The experiment ran for a fixed number of generations, and in twelve out of the eighteen total trials, ancestor B predominated in the population. While this result does not appear to overwhelmingly support the quasispecies model, each of the six trials in which ancestor A dominated had such small differences between the two ancestors that the results were within the expected experimental error.

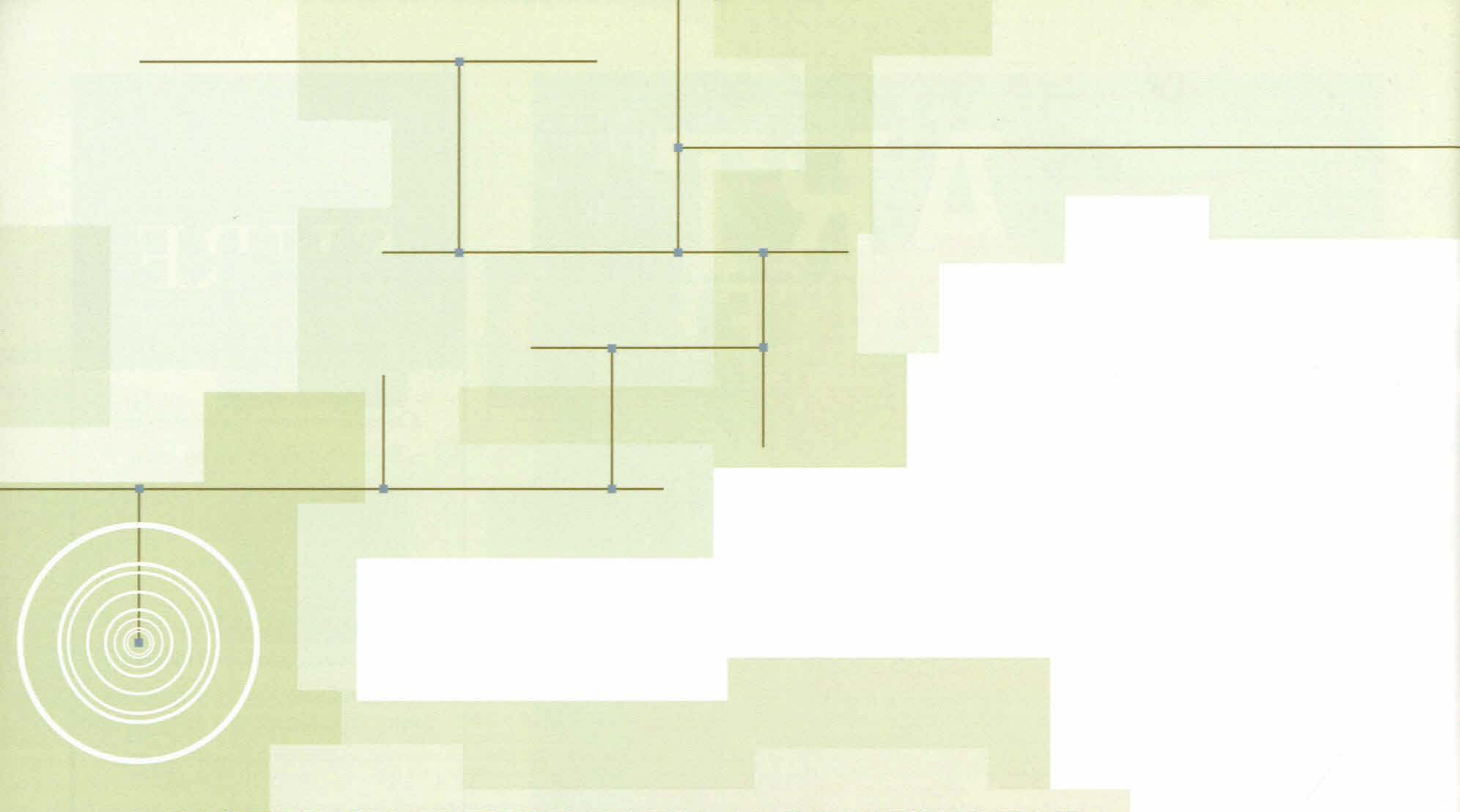
These initial results suggest that replication rate is not the only factor that governs the success of a species. Ancestor B's ability to match, and in many cases, dominate over ancestor A in terms of numbers shows that robustness is key to the population's strength. This study compels researchers to compare the dynamics of digital and biochemical organisms and to reevaluate previous notions of genetic fitness.

The next step is to determine if this new theory also holds for biological organisms. *E. coli* would be a suitable test subject due to its simple asexual reproduction that is well modeled by digital organisms. Perhaps then, we may find that evolution depends not on the survival of the fittest, but the survival of the flattest replication peak. ■

Jialan Wang is a junior in Mathematics at the California Institute of Technology. *She wishes to thank her mentors Chris Adami and Claus Wilke for their support on this project.*

FURTHER READING

1. C. O. Wilke, J. L. Wang, C. Ofria, R. E. Lenski, and C. Adami. Evolution of digital organisms at high mutation rate leads to survival of the flat test. *Nature* **412**, 331-333 (2001).
2. C. Adami. *Introduction to Artificial Life* (S. Verlag, New York, NY, 1998).
3. M. Eigen and P. Schuster. *The Hypercycle: A principle of Natural Self-Organization.* (Springer, Berlin, 1979).



A FASTER SHORTEST-PATH ALGORITHM FOR MANHATTAN SPACE

BY PO-SHEN LOH

YOU'RE AT AN OUTDOOR CAFÉ IN DOWNTOWN MANHATTAN when your cell phone rings. It's the Department of Finance telling you that your car will be towed in thirty minutes. Your brain snaps into action, mentally traversing the long, narrow streets and considering which way to turn at each intersection. If you're smart, you'll also consider the fact that no subway construction is fencing in your car on three sides. How can you figure out the shortest path to your car? More importantly, can you do it quickly enough?

These aren't just questions for absent-minded motorists. The pathways on circuit boards and utility lines follow the same pattern of connections: every path must be parallel to an axis (see Figure 1). F. O. Hadlock was one of the first to tackle this problem, devising a method that begins at a single point and radiates outward like the waves of a pebble thrown into a still pond. This so-called wavefront expands with each iteration and adds adjacent nodes to itself one by one. For each node it adds, it creates an estimate of the shortest path that goes through that node. Eventually the wavefront will reach the destination, and the estimate will now be the actual length of the shortest path in what mathematicians call "Manhattan Space."

"The pathways on circuit boards as well as utility lines follow the same **pattern of connections...**"

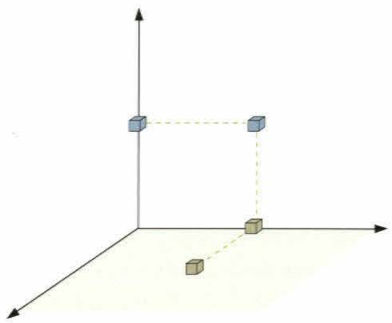


FIGURE 1. Manhattan Space: every path is parallel to an axis.

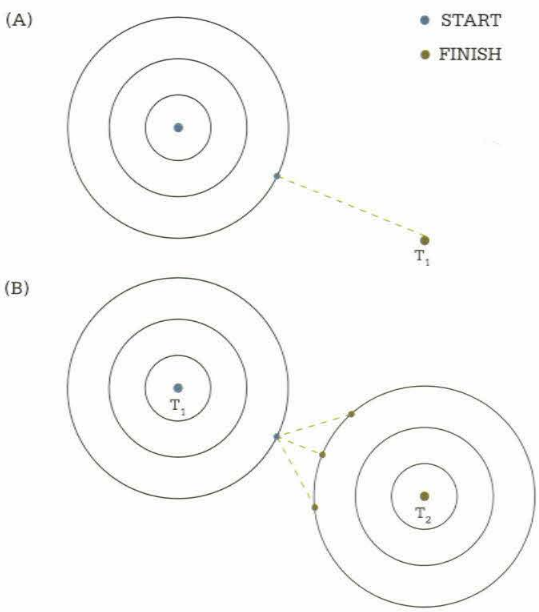


FIGURE 2. The two types of wavefronts: (A) Front-to-end searches only consider the destination when computing the growth of the wavefront; (B) Front-to-Front searches consider points on the other wavefront when computing the growth of the wavefront.

These front-to-end algorithms have no information about the conditions around the destination terminal. As the search hunts through intersection after intersection searching for the shortest path, a front-to-end algorithm might waste time and effort going down seemingly promising blind alleys. The alternative is to send wavefronts from both terminals rather than from just one. As these front-to-front algorithms search from both directions, the wavefronts cooperatively move toward each other until they merge and the shortest path is found.

Despite the intuitive nature of front-to-front searches, they naturally require a greater overhead in terms of computational complexity, which measures approximately the maximum time a given procedure would take within a constant factor. The computer science community sees these worst-case front-to-front searches as so ungainly that front-to-end searches have become the de facto standard. The complexity functions of both front-to-front and front-to-end searches are parameterized by the number of nodes examined by the search; therefore, it is advantageous to develop an algorithm that reduces the number of nodes, even if doing so incurs a small additional complexity factor per node.

We present a new front-to-front algorithm for shortest paths in Manhattan space that adds only a nominal log-power complexity cost per node, yet can reduce the number of examined nodes by an n^{th} root in common situations. We can have a reasonable worst-case time scenario while still reaping best-case advantages of a front-to-front search.

DIMENSIONAL AUGMENTATION AND THE CLOSEST POINT INTERPRETATION

Because the overhead cost of front-to-front searches stems from finding pairs of wavefront points that strictly minimize the total length, we propose an algorithm that relaxes the minimality restraint without compromising validity. The first step is to reduce the problem of estimating the total length to a problem of selecting two minimally-separated points from two separated sets.

To accomplish this, we enter $\mathbb{R}^{d+1}$ space–that is, we add another dimension to the problem by embedding the wavefront points according the following criteria: we define σ_k as the length of the shortest path joining an intersection adjacent to a point p in wavefront W_k. Because we will frequently move between $\mathbb{R}^d$ and $\mathbb{R}^{d+1}$ space in the course of this discussion, we will employ the following notational convention: if a point in $\mathbb{R}^{d+1}$ is marked with a "prime" (P') the un-primed point (P) will denote its analog in $\mathbb{R}^d$.

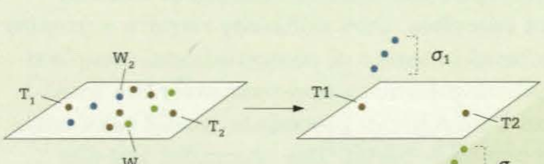


FIGURE 3. Transformation of our wavefronts into $\mathbb{R}^{d+1}$ space. The estimated length is now the distance between both points.

Dimensional augmentation simplifies the estimation of the total path length, as the estimated length of the shortest path is now just the Manhattan distance between the two points in $\mathbb{R}^{d+1}$. We have made the problem of estimating the shortest path an exercise in measurement. This problem is known as the Manhattan dynamic bichromatic closest-pair problem and has already been solved efficiently; there are algorithms that solve it in $\log^{d+2} n$ cases for n points in d dimensions. For our purposes, it suffices to find "almost-closest" pairs, which in fact speeds up our algorithm by a factor of $\log n$, providing a considerable boost for lower dimensions.

For the purposes of expediting the explanation of the algorithm, we compartmentalize these computations within a Dynamic Closest Point Structure (DCPS), which holds the wavefront points in $\mathbb{R}^{d+1}$. We construct the structures so that we can add and delete points as well as perform a query operation on a point. This query operation takes a point P' in $\mathbb{R}^{d+1}$ that is not contained in the DCPS and returns the point in the DCPS that is closest to P'. Now that we are armed with this heuristic, it remains only to establish the termination condition. The search heuristic for each point P underestimates the true length of the shortest path between the terminals that contains P; hence if we examine the points in order of their search heuristic, we can stop as soon as the value of the heuristic exceeds the length of a discovered path.

PULLING THE PIECES INTO AN ALGORITHM

We're now almost ready to talk about how the algorithm works in connecting a start point T_1 to a final point T_2, but first we need to define a tool that we'll use along the way. A priority queue contains a set of points with two associated pieces of information: the estimated length of the shortest path containing that point and an estimate of how close the point is to the second wavefront. When we remove an item that we have added to a queue, this "pop" instruction removes the item with the shortest estimated path. If there is a tie between shortest paths, then we choose the element closest to the second wavefront, thus ushering the wavefronts toward each other.

The wavefronts begin by containing just two endpoints in $\mathbb{R}^{d+1}$. Right now, we can only have one candidate path–directly from T_1 to T_2–which is placed into the priority queue with the distance from T_1' to T_2' as the estimated path length and the distance from T_1 to T_2 in $\mathbb{R}^d$ as the estimated distance to the second wavefront, which is now just the final destination. We also set the variable MIN to

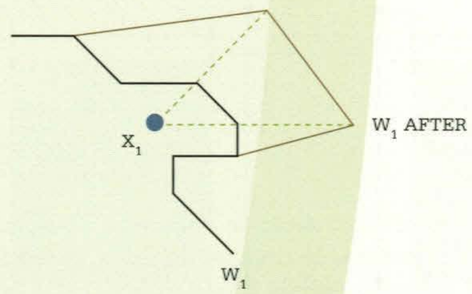


FIGURE 4. The growth of the wavefront as it considers each adjacent point V.

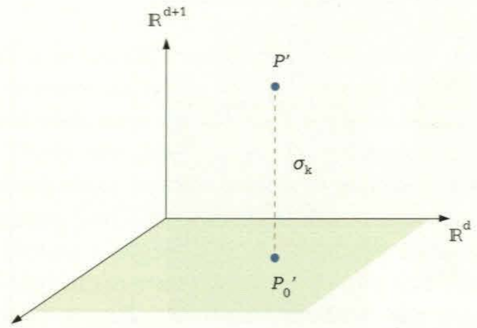


FIGURE 5. Moving a point back into the plane of the problem.

infinity–this variable represents the length of the shortest path found so far, and since we have not yet found any paths between the terminals, it starts with an infinite value.

We now enter into the main loop of the program which repeats until the queue or the wavefronts (W_1 and W_2) in $\mathbb{R}^{d+1}$ are empty. The first step of the loop is to pop off the minimum element that is currently in the queue and also in W_1; we throw out any points we come across before we get to a point in the wavefront. We now check to see if the estimated length for this point, P', is greater than MIN. When it is, we have a shortest path because the estimated length is an underestimate of the true length of the shortest path through each point. MIN is the length of an actual path between the terminals, and since P' was the point that got popped off the priority queue, all other points on the priority queue yield lengths greater than MIN. Therefore, we can short-circuit the rest of the algorithm and terminate because MIN must correspond to the shortest path.

If we don't terminate, we define a new point, P_0', that is the projection of P' into the plane with zero as its $(d+1)$-st coordinate. We now want to expand our wavefronts, so we define X_1' to be the closest point to P_0' in W_1' and X_2' to be the closest point to X_1' in W_2'. Next we check to see if our wavefronts have crossed.

If X_1' is in the visited points of the second wavefront, we restart the loop. Otherwise, we remove X_1' from its wavefront and add it to the set of examined points of W_1. Then for every V adjacent to X_1, we embed V into $\mathbb{R}^{d+1}$ space by adding σ_k as the $(d+1)$-st term when P is in W_1, σ_k when P is in W_2, and 0 when P is in T_1 or T_2. We add V' to the wavefront and put it into the priority queue. We do the same for second wavefront, except we do not add points into the priority queue. Repeat until the termination conditions are met.

BOUNDING THE NUMBER OF ITERATIONS

One of the defining characteristics of any algorithm is its time complexity. No matter how clever or thorough an algorithm is, it is worthless if it cannot produce results in a reasonable amount of time. To compute time complexity, we have to bound the number of operations our algorithm requires. We first bound the number of times the loop can cycle.

"Dimensional augmentation simplifies the estimation of the total path length, as the estimated length of the shortest path is now just the Manhattan distance between the two points in $\mathbb{R}^{d+1}$."

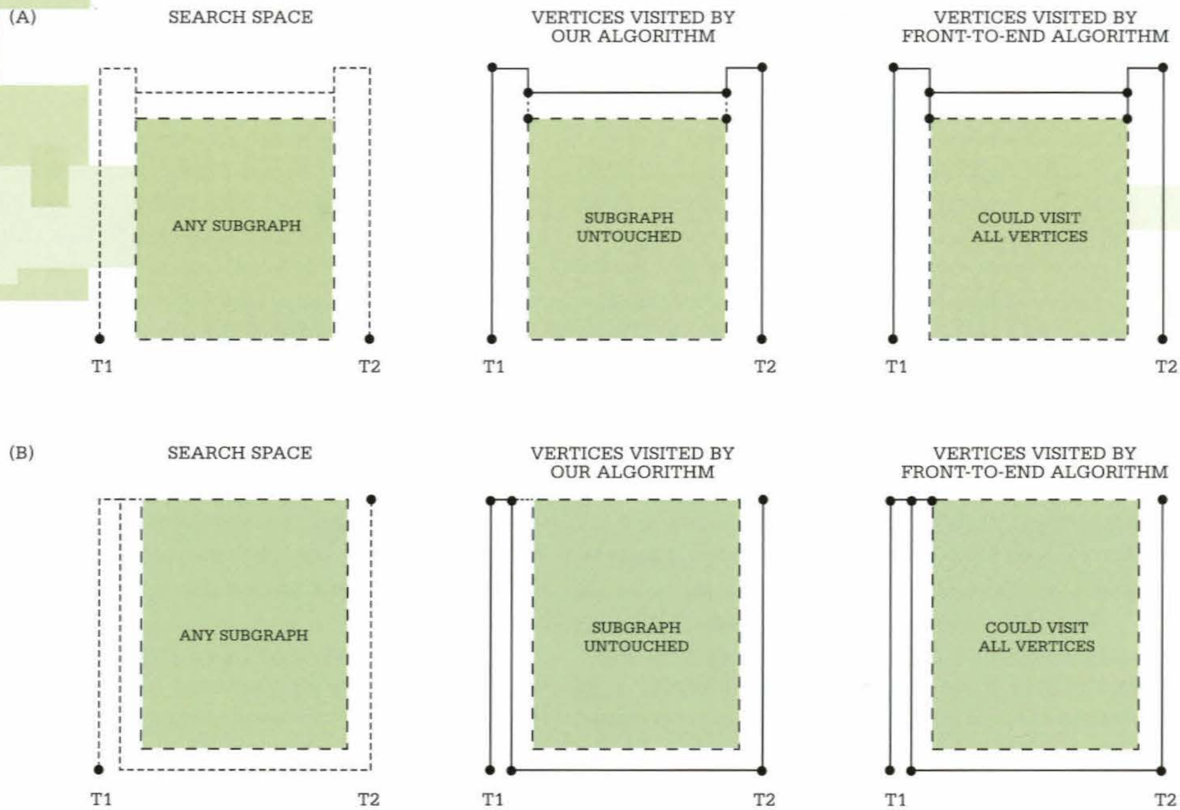


FIGURE 6. (A) The subgraph can be arbitrarily complex, but our alogorithm will always visit 8 vertices while the Front-to-end algorithm must visit all subgraph vertices with priority less than T1 T2. Those vertices include all points that can be reached from the top-left vertex by moving down and to the right along the edges in the graph. (B) is another family of search spaces in which our algorithm always visits a fixed number of vertices while the front-to-end can visit arbitrarily many.

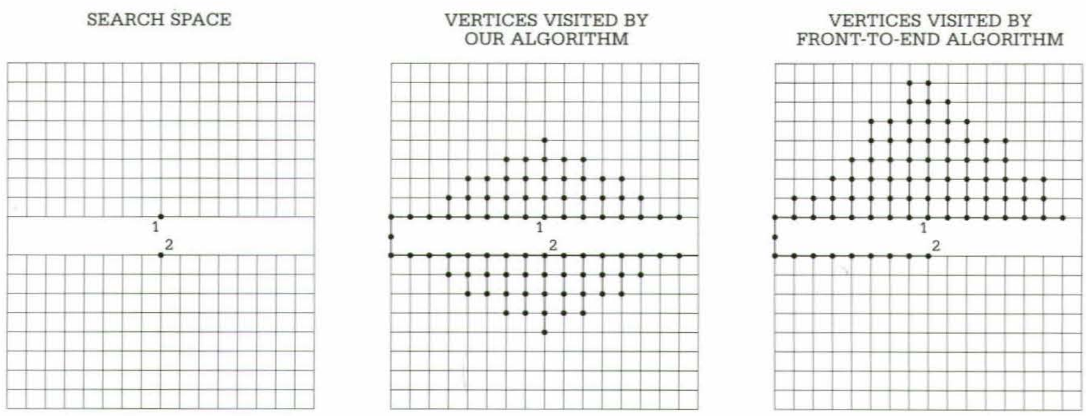


FIGURE 7. A plain-vanilla case: the vertices marked 1 and 2 are the respective terminals. Our algorithm rarely hits the bound of the theory, even when the search space does not clearly favor bidirectionality. In this particular case, our algorithm and the front-to-end alogorithm visit about the same number of vertices.

The main loop runs at most $(2d+1)n+1$ times, where d is the dimension of the search space and n is the number of visited vertices. By the nature of our space, each point can only have $2d$ incident paths since there are d axes and two directions an incident path could approach a point. There are three possible cases for each time the loop goes through: it can add a point in the visited nodes and insert up to $2d$ priority queue elements, it can reduce the number of such elements by one, or it can be the final loop.

All we have to do is find a bound for the number of times each type of loop can occur. Since the number of visited nodes is bounded by the total number of nodes, σ_1 can grow no more than n times, so we can have at most n of the first iteration type. The number of insertions into the queue is then bounded by $2dn$ since there are n points with a maximum of $2d$ neighbors. And since the number of iterations of the third type is obviously one, we have a total of $(2d+1)^d + 1$ loops.

The worst-case time complexity of our algorithm is $n (\log n)^d$. By far the most computationally intensive method within the loop is a query or update of our DCPS. This can be completed in $(\log n)^d$ time, so we now multiply the number of times this happens by the number of loops we complete. Dropping the proportionality to constants, we have a time complexity of $n (\log n)^d$.

THE REAL WORLD: TESTS AND APPLICATIONS

These theoretical bounds are all upper limits; our algorithm usually performs much better than these worst-case bounds. There are many cases where a traditional front-to-end algorithm takes much longer. Figures 6(A) and (B) present a couple of cases where our algorithm would visit a relatively small number of vertices while a front-to-end algorithm would visit every node in the green subgraph.

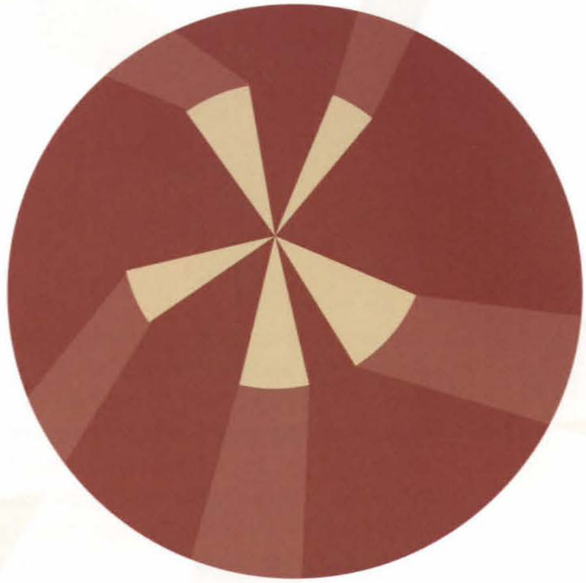
There are also many cases in which a problem does not clearly favor bidirectional searches, but our algorithm still remains useful. Figure 7 presents a case where a front-to-end algorithm visits approximately the same number of nodes as our algorithm. We therefore conclude that while this algorithm is at a slight disadvantage in the worst case, there are many applications in electronics design or layout where the use of a bi-directional algorithm would allow a quick determination of the optimal path.

So what advice do we have for a poor soul racing against the clock and the Department of Finance in such a rectilinear world? Head for the nearest elevator; breaking out of the 2D plane of your problem will give you a new perspective on your path. C

Po-Shen Loh is a third year undergraduate in Math at the California Institute of Technology. *This research was supported by Axline and Larson fellowships, and was conducted under the mentorship of Alain Martin with additional assistance from Po-Ru Loh, Mika Nyström, and Charles Leiserson.*

FURTHER READING

1. D. Eppstein. Dynamic Euclidean Minimum Spanning Trees and Extrema of Binary Functions. *Discrete Computational Geometry* **13**, 11-122 (1995).
2. F. O. Hadlock. A Shortest path Algorithm for Grid Graphs. *Networks* **7**, 333-334 (1977).
3. H. Kaindl and G. Kainz. Bidirectional Heuristic Search Reconsidered. *Journal of Artificial Intelligence Research* **7**, 283-317 (1997).



LENSES TO HARVEST THE SUN'S ENERGY

BY JANET Q. ZHOU

CALIFORNIA'S ENERGY CRISIS DURING THE SUMMER OF 2001 EMPHASIZED THE NEED TO develop alternative energy sources. Previous efforts to create highly efficient solar cells have made little progress, leaving us to wonder what will happen when our existing supply of fossil fuels runs out. New types of lenses may provide the key to harvesting the sun's energy.

Visionaries have long dreamt of solar-powered homes, vehicles, and electronics. However, existing solar technology is both inefficient and expensive at the present time, making it unlikely to replace fossil fuels as our primary source of power any time in the near future. With the Earth's supply of fossil fuels slowly dwindling and expected to expire sometime within the next century, solar energy has emerged as an attractive alternative because it is free, clean, and renewable. However, the problem lies in inefficient photovoltaic systems. We can increase the efficiency of solar cells by designing new lenses that are able to collect more sunlight. Affordable and practical solar power is fast becoming a reality.

"Previous efforts to create highly efficient solar cells have made little progress, leaving us to wonder what will happen when our existing supply of fossil fuels runs out."

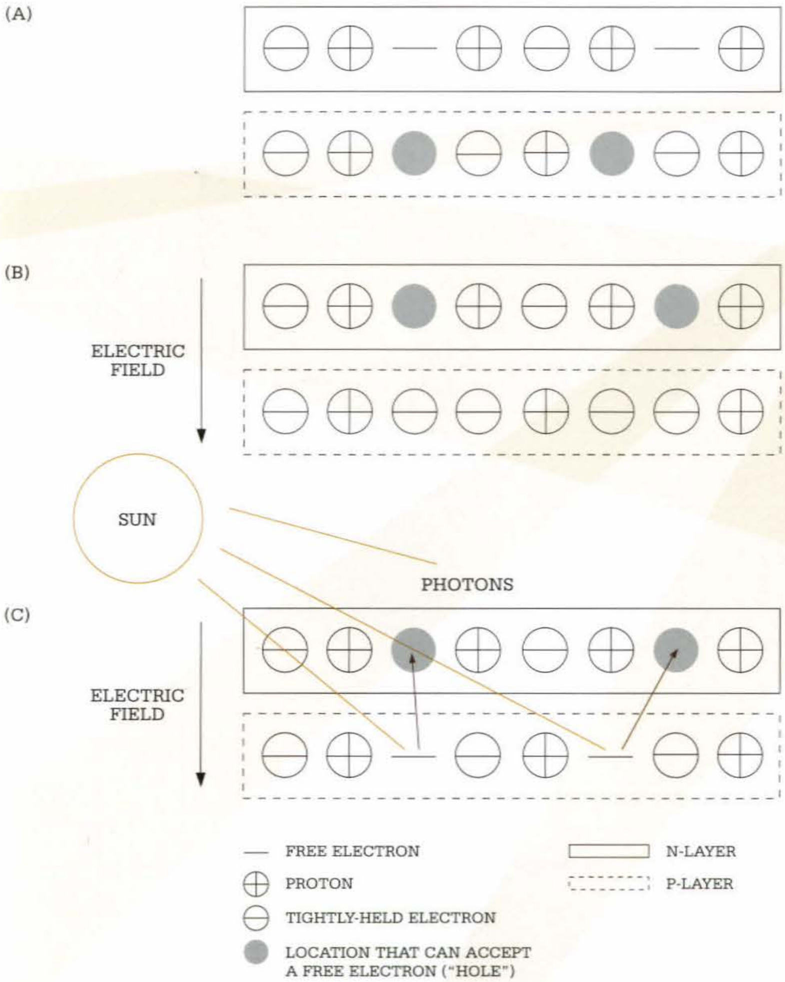


FIGURE 1. How a photovoltaic cell works. (A) Excess electrons in the front half of the cell flow towards the back half. Positive charges (holes) form where the electrons leave their positions. (B) The imbalance of charge causes an electric field to form across the junction which prevents any further crossover. (C) Photons hitting the semiconductor knock electrons loose. The electrons are pushed toward the front surface of the cell by the electric field, generating a current, which can power an external load.

FROM SUNLIGHT TO ELECTRICITY

Photovoltaic cells (see Figure 1), also known as solar cells, have been used to electrically power devices as small as watches and hand held calculators to machinery as large as water pumps and communications systems. Composed of semiconducting materials such as silicon, these photovoltaic cells convert light into electricity using the photoelectric effect–photons hitting the semiconductor knock electrons loose, allowing the electrons to flow through the material, thus producing an electric current.

The amount of electric current produced helps to determine the cell's efficiency. Only sunlight of particular energies will effectively create electricity, and the materials that make up the cell often reflect or absorb most of the sunlight. Typical commercial photovoltaic cells currently have efficiencies ranging from about eight to fifteen percent, that is, roughly one-sixth of the sunlight striking the cell generates electricity. Low efficiencies mean that larger arrays are needed, resulting in higher costs. Efficiency increases with increased sunlight absorption.

Certain photovoltaic cells may be built into concentrating arrays that use lenses to focus the absorbed sunlight. The idea is to use as little of the expensive semiconducting material as possible while collecting the maximum amount of sunlight. The main components of the design include a thin film solar cell, an array of soft polymer microlenses, and a heat sink to conduct heat from the system.

DIFFERENT LENSES TO SOAK UP THE SUN

Higher efficiencies require lenses that best concentrate the absorbed sunlight. The two different types of lenses used were solid immersion lenses and fresnel lenses. The design consisted of a fresnel lens atop two solid immersion lenses to focus the incoming light to 40 to 100 times the sun's average energy. This concentration is sufficient to significantly increase the efficiency of solar cells. Solid immersion lenses (see Figure 2) are greater than hemispherical. Imagine that they are the large parts of spheres cut along portions other than their diameters. Light from any direction will focus at the center of the base of the solid immersion lens where the substrate, a thin film cell, is placed.

A fresnel lens (see Figure 3) performs the same function as a conventional lens, but it has the advantage of being much thinner and of weighing much less. The lens consists of many small concentric rings with jagged edges, each focusing the light toward a central focal point. We remove the portion of the conventional lens that does not contribute to the focusing power of the lens, leaving only the top layer, which forms the fresnel lens.

We made the lenses out of polydimethylsiloxane, a room temperature vulcanized rubber compound that is inexpensive and easy to mold. It is well known for its strength and durability to shock, moisture, and other environmental hazards. This material is ideal for optical use since it is clear and has a refractive index that is very close to that of glass. Significant advantages of polydiethylsiloxane are that it allows the production of interesting shapes that are unattainable in glass and that it also retains its properties up to 200 degrees Celsius, which is important since a large amount of heat will flow through the lenses.

To fabricate the solid immersion lens, we first pour a small amount of material into a petri dish of known area. (see Figure 4). We then place six ruby ball bearings spaced evenly into the dish. After the material cures,

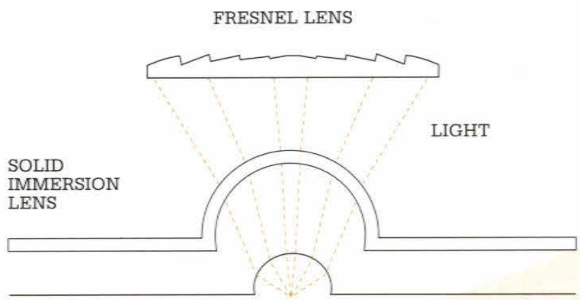


FIGURE 2. Schematic diagram of a solid immersion lens.

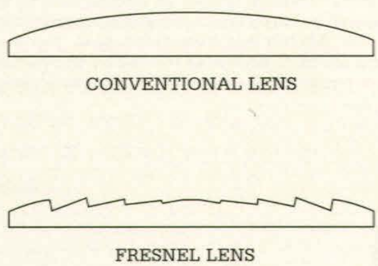


FIGURE 3. Comparison of a fresnel lens to a conventional lens, in which superfluous portions are removed. The fresnel lens helps to focus light towards a central point.

"Solid immersion lenses are greater than hemispherical. Imagine that they are the large parts of spheres..."

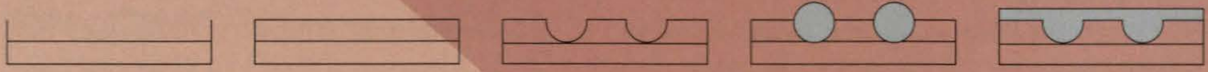


FIGURE 4. Fabrication process of solid immersion lenses, using molds within a petri dish. Molds are formed using ball bearings and can be reused as often as required.

FIGURE 5. Removing a fresnel lens from its mold. The fresnel lens can be molded from an existing glass lens.

we remove the ball bearings with a pair of tweezers. We calculate the mass of polydimethylsiloxane needed to fabricate the lens and pour this amount into the dish. After degassing and placing the samples in the oven, we peel the lenses from their molds using a pair of tweezers. We are able to reuse the molds as often as needed and can make solid immersion lenses on the order of millimeters to centimeters in radius. The only drawback is that we cannot accurately control the thickness of a layer when trying to make smaller lenses.

Fresnel lenses are easier to fabricate than solid immersion lenses since we can use existing glass lenses to make them. We make a mold of the lens by first pouring a layer of polydimethylsiloxane into a petri dish that is larger than the size desired. We then degass and cure the first layer. Then we place the glass lens onto the material and apply gentle pressure until all parts of the lens stick to the material. We cover the lens completely with an excess of polydimethylsiloxane to weigh it down and after degassing and curing, we remove the lens from the molds by peeling the layers apart (see Figure 5).

TRANSMISSION WITH MINIMAL LOSS

We made nine samples to determine the transmittance of the material. The samples varied among three different widths and three different compositions of polydimethylsiloxane. Using a Fourier transform infrared (FTIR) spectrometer, we measured the transmittances of the samples of three varying compositions against a background of air.

There are common features such as strong absorption below 4500 wavenumbers. A summary of the major absorption features is presented in Figure 6. Since most of the light that was converted was in the visual range, transmittance spectrums in the 550 nm to 1000 nm range were taken using an optical multichannel analyzer (OMA).

In the 1.25 mm width group, the three samples have transmittances that are between 90 percent and 95 percent. For the next group, the three samples have transmittances that are further apart, ranging from 91 percent to 98 percent. The 5.0 mm width group exhibits transmittances that range from 80 percent to nearly 100 percent. This trend is expected since thicker lenses result in greater absorption, which lowers transmittance.

We examined and photographed the fabricated lenses under a microscope to determine the types and levels of acceptable aberrations that could potentially affect the performance of the lenses. Additionally, we characterized the focal lengths and magnifications of the lenses. The focal lengths were examined using a setup that held the lens in place while subjecting it to light. We moved the lens until the size of the spot was the smallest and most focused and estimated magnification using an optical microscope. We placed a calibration sample under the microscope (see Figure 8(A))

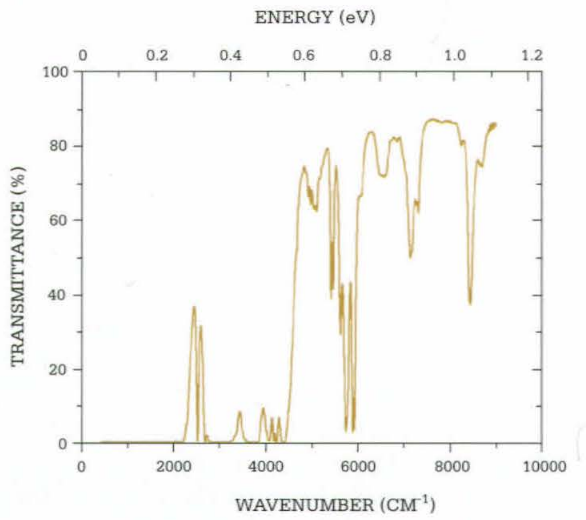


FIGURE 6. The graph of transmittance versus wavenumber seen here is for one of the samples. There were no noticeable differences among the three samples.

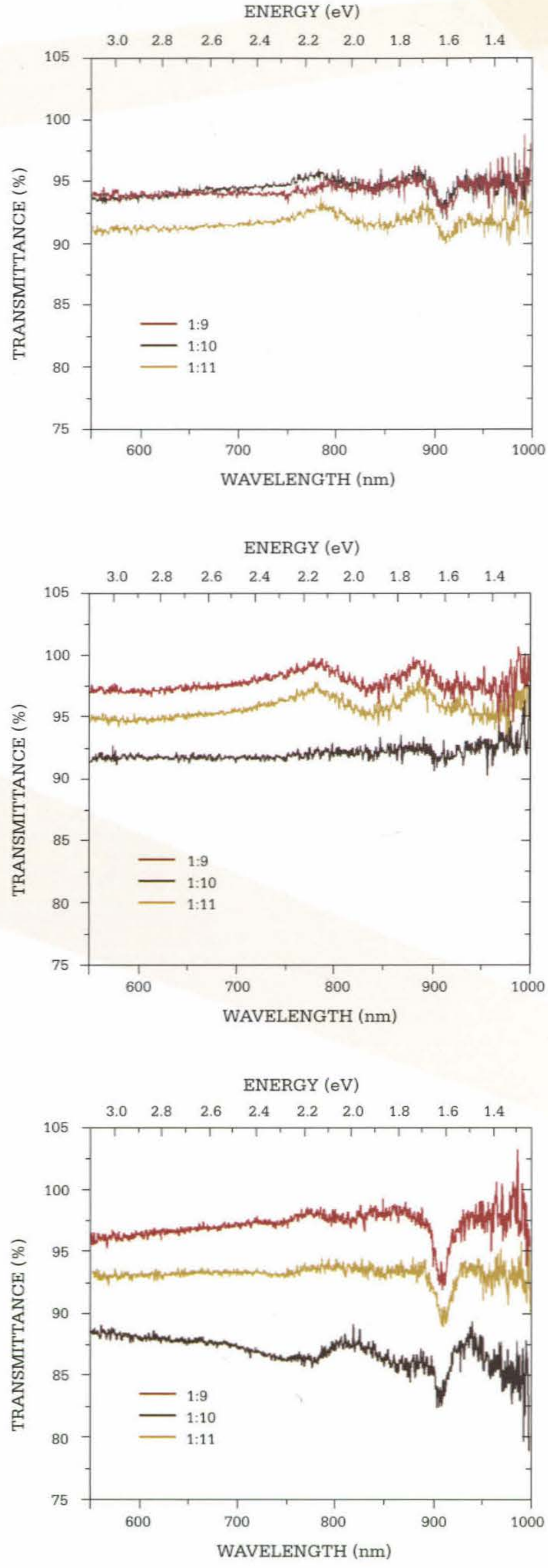


FIGURE 7. Transmittance spectrums for various samples indicating transmission between 80 and 100%. From top to bottom, the graphs are those of groups with widths 1.25mm, 2.5mm, and 5.0mm.

and the widths of the light and dark bands were measured and then compared to the widths of the bands when a lens was placed above (see Figure 8(B)).

The magnifications of solid immersion lenses and double solid immersion lenses agreed with expectations. The solid immersion lenses magnified images by a factor of about two, and the combination of a double solid immersion lens and a solid immersion lens had a magnification of four, since each solid immersion lens contributed a factor of two to the final magnification. The magnifications of the fresnel lenses were smaller than expected, but this may have been due to difficulties in focusing the images under the lenses or due to poor magnification by existing lenses.

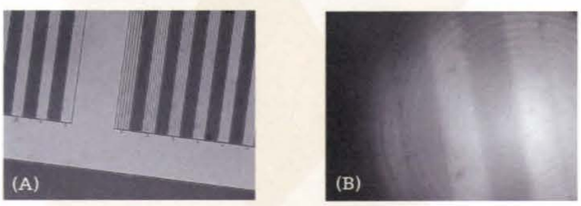


FIGURE 8. (A) is a calibration sample under microscope. (B) is the sample with a lens placed on top.

THE FUTURE OF SOLAR TECHNOLOGY

The experiment concludes with the successful fabrication of a lens design that collects sunlight at a higher efficiency than ever before. Results indicating effciencies mostly near 100 percent are promising. The properties of polydimethylsiloxane were desirable, being strong and pliable while retaining optical properties similar to that of glass. Variability in transmittance increases when the width of the lens is increased. With greater widths, the material absorbs more of the incoming light, causing transmittance to decrease. Results show that the microlens system will efficiently and affordably transmit and focus incoming light. It can effectively duplicate and possibly expand the functions of glass lenses.

An optical program is currently being used to determine the optimal size for the fabrication of new lenses. Additional aspects of the design will then be integrated to form a cell, which will be tested for power and efficiency. The result of this project will be a highly efficient photovoltaic cell that will help meet the challenge of low-cost solar cells.

Janet Zhou is a third year undergraduate in Electrical Engineering at the California Institute of Technology. *The work was completed with Professor Harry A. Atwater, Professor of Applied Physics at Caltech, and Jimmy Zahler, the graduate student overseeing this work. The work was funded by the 2001 Arthur E. Lamel Memorial Summer Undergraduate Research Fellowship.*

FURTHER READING

1 Swanson, Richard M. Promise of Concentrators. *Progress in Photovoltaics: Research and Applications.* **8**, 93-111 (2000).

2 Winston, Roland. Nonimaging Optics. *Scientific American* **264**, 76-81 (1991).

3 Wu, Qiang, et al. Realization of numerical aperture 2.0 using a gallium phosphide solid immersion lens. *Applied Physics Letters* **75**, (1999).

STEALING SECRETS BY JAIME ZOLLARS

Stealing Secrets
mixed media, 8" x 10.5"
Artist's Private Collection

In the near future, beating a lie detector test will be out of the question. Scientists are looking deep within the brain to root out deception at its source. Machines and behaviorists can work together to discover even the smallest white lie while also determining whether the lie was premeditated or spontaneous. Technology and biology come together in fascinating and often disturbing ways.

Jaime Zollars
Art Center College of Design