

*Medal
of
Honor*

ANDREW VITERBI'S
Fabulous Formula

HIS DISCOVERY LED TO 3G CELLPHONES, WI-FI,
AND A HOST OF OTHER TECHNOLOGIES

BY
TEKLA S.
PERRY

THE RANKS OF FIRST-RATE INVENTORS are chock-full of characters who are brash, egotistical, and temperamental. And for quite a few, even those adjectives are stretched to their euphemistic limits.

So meeting Andrew J. Viterbi can be something of a shock. He speaks softly, responds patiently. It's not that he's shy; it's a soothing kind of quiet, the kind that makes a guest comfortable. He's dressed nicely—in gray slacks and a dress shirt—but not formally; he rarely wears a tie. He's mostly bald, with a round face and eyes that crinkle when he smiles, which is frequently. He looks like someone's grandfather (which indeed he is; he has five grandchildren).

47 NA
IEEE SPECTRUM
MAY 2010

"Success comes to all kinds of personalities," says Viterbi's friend Carver Mead, the Caltech professor, integrated circuit pioneer, and oft-quoted tech visionary. "But it sure is nice when it comes to people like Andrew, who aren't just in it to beat their own chests."

It's not easy to reconcile the mild-mannered Viterbi with the tech titan who made fundamental contributions to Wi-Fi, 3G cellular and digital-satellite communications, speech recognition, and DNA analysis. Who cofounded Qualcomm. And, oh yeah, came up with one of the most important mathematical concepts of the 20th century: the Viterbi algorithm, a means of separating information from background noise. It's that last one, the algorithm, that was singled out in the citation for the IEEE Medal of Honor, Viterbi's most recent accolade.

Viterbi's tale isn't one of an aggressively ambitious engineer trying to shake up the world, make a fortune, or claw his way to the top of his profession. It's the story of an unusually bright and hardworking professor who wanted to explain a difficult concept in clear and simple terms in order to better teach his students.

THE SON OF JEWISH-ITALIAN immigrants, Viterbi did well in both math and English at the venerable Boston Latin School. His father, a doctor, encouraged him to be an engineer, remembering the impact electrical power had made when it first came to Bergamo, the Italian town where Andrew was born. Viterbi won a scholarship to MIT and graduated with bachelor's and master's degrees in electrical engineering in 1957. His father's medical practice was struggling, and the family needed Viterbi's financial support, although he wanted to go on to a Ph.D. and teach.

He had enjoyed working at Raytheon as a co-op student, but deplored the way engineers were treated. "Engineers were not people trusted to make any decisions," he recalls. He'd heard that things were different on the West Coast, where some engineers even got private offices. So he applied for and got a job at the Jet Propulsion Laboratory, in Pasadena, Calif. Signing on with a lab that was affiliated with a university seemed like the next best thing to the academic career he really wanted.

At JPL, he started off working on telemetry for guided missiles, helping develop a then-new device called the phase-lock loop, which tunes into a carrier signal in spite of surrounding noise. After the Soviet launch of Sputnik and the beginning of the space race, Viterbi's efforts shifted to space communications systems, but the underlying focus on signals and noise didn't change. Simultaneously he worked on his Ph.D. at the University of Southern California.



Andrew J. Viterbi

DATE OF BIRTH

9 March 1935

BIRTHPLACE

Bergamo, Italy

FIRST JOB

Soda jerk in a drugstore

FIRST JOB IN TECHNOLOGY

Co-op student, Raytheon

PATENTS

Six, but inspired hundreds

HERO

Claude Shannon

MOST RECENT BOOK READ

Skeletons at the Feast by Christopher A. Bohjalian

FAVORITE MUSIC

Light classical, opera, music of the '40s and '50s

COMPUTER

Lenovo laptop, several Dell

FAVORITE RESTAURANT

Il Tinello, New York City

FAVORITE MOVIE

Casablanca

BIGGEST WORRY

Nuclear proliferation

LANGUAGES SPOKEN

English, Italian, French, German

ORGANIZATIONAL MEMBERSHIPS

IEEE, National Academy of Sciences, National Academy of Engineering, Jewish Community Foundation of San Diego, Mathematical Sciences Research Institute

In the fall of 1963, doctorate in hand, Viterbi finally made it to academia, joining the University of California, Los Angeles. Teaching classes in communications and information theory, he couldn't have been happier.

Then came the algorithm.

IT WAS MARCH 1966. Viterbi was struggling with yet another class of graduate students, who just couldn't grasp a key set of concepts in information theory. The troublesome algorithms, known as sequential decoding of convolutional codes, extracted data from a signal corrupted with noise. Essentially, the algorithms decided if a bit was a 0 or a 1 by looking down a decision tree. When it became clear that the data had been corrupted, the software would go back one or more steps and try a different path.

The students didn't get it. Viterbi decided that the reason the algorithms were so hard to understand was that the proof of the theorems was too complex. So he set out to find a simpler proof.

After several months of obsessing over the problem, it hit him: It wasn't the proof that needed to be simplified; it was the algorithms themselves. Instead of going down a tree over and over again, Viterbi envisioned a trellis, in which the software considers the bits surrounding a particular bit in question to decide whether that bit is a 0 or a 1. The software assigns a probability of the accuracy of its decision to each bit based on the voltage of the received signal that conveys that bit. Based on the probabilities, the algorithm then decides whether that particular bit is a 0 or a 1. Unlike the earlier convolutional code algorithms, the software needs to keep track of only the paths leading up to a limited number of states, typically more than four but not more than 1000, a far

more effective method than following each path until it dead-ends. Viterbi published his results in the *IEEE Transactions on Information Theory* in 1967, and his paper became a classic.

"After you see this approach, you wonder why nobody thought of it before," says Robert G. Gallager, an MIT professor emeritus and an eminent scholar in communications theory. "But that's what the best inventions are. After you see them, they are obvious."

The algorithm did what Viterbi wanted—it simplified the course material for his students. But he sensed it could do a lot more—for example, enabling the extraction of weaker signals from noisier environments. That in turn could mean lower-power transmitters, smaller antennas on the receiving end, or both. But to be useful, it would require both computer memory and processing power to calculate and track all the probabilities. Extracting the weakest signal from the greatest amount of noise would mean

BRAD SWONETZ

keeping track of about 1000 states at once; to do that, you'd need the processing power of a mainframe computer.

Viterbi and his colleagues did some further tinkering with the algorithm. They discovered that by keeping track of just 64 states, you could create a device that was four times as good as an uncoded transmission, or twice as good as coding systems in use at that time. That meant that the transmission power could be one-fourth the strength, or the receiving dish one-fourth the size, or the range twice as far, as similar uncoded systems. Within a few years, the falling price of electronic components made possible devices that tracked 256 states.

In 1968 Viterbi joined two engineers from his JPL days—Irwin Jacobs, then at the University of California, San Diego, and Leonard Kleinrock, then at UCLA—and started consulting on applications of his algorithm. They called their firm Linkabit.

The company worked on various defense and commercial satellite modems and terminals. It also built a satellite signal scrambler called Videocipher for the cable channel HBO; the technology continued to be used to scramble pay-TV signals until the end of 2008. In 1973, Viterbi left UCLA and joined Linkabit full time. Seven years later, M/A-Com acquired the company and eventually sold off its various pieces. Viterbi and Jacobs stuck around until 1985, when they decided to start all over with a new business they named Qualcomm. They weren't exactly sure what this company would do, just that it would be something in commercial communications.

Then, recalls Viterbi, "along came this interesting fellow, Allen Salmasi." With backing from a rich uncle in Paris, the Iranian émigré had founded a company called OmniNet. Salmasi envisioned a mobile satellite network that would let trucking dispatchers track trucks in real time and send messages to the drivers. He hired Qualcomm to build it.

The time was right. A number of companies had launched satellite TV businesses, but the service wasn't catching on, so they were eager to lease their transponders. There was just one problem: Those satellite downlinks were licensed for fixed reception, not mobile use. The only way around the rule was if the mobile application did not interfere with fixed services.

Qualcomm's solution was to use spread-spectrum communications. The engineers figured that if they combined their signal with a broader signal that looked like noise, the fixed satellite networks would ignore it. They then used the Viterbi algorithm to help extract the original signal from the noise. The Federal Communications Commission gave

Qualcomm an experimental license to try out the idea.

In 1988, Qualcomm was in the midst of testing the system with 600 trucks when Salmasi's company started falling apart. Viterbi and his colleagues, rather than letting the effort fold, decided to acquire OmniNet in 1988. Within three years, the trucking system, called OmniTracs, was turning a profit. It's still used around the world by long-haul truckers.

FOLLOWING THE LAUNCH OF OMNITRACS, the scrappy start-up took on the mobile phone industry. In an era when analog cell-phones still ruled, the company introduced a digital spread-spectrum network (also based on the Viterbi algorithm, of

PENN STATE | ONLINE



Online Master's Degree in Systems Engineering

Advance Your Career

- Gain a quality education in a convenient online format
- Apply your skills to any engineering discipline
- Build a professional network with classmates
- Become a leader in your organization
- Finish in as little as two years



Apply now

www.worldcampus.psu.edu/IEEE

U.Ed.OUT 10-0388/10-WC-151bkh/bjm

SPECTRUM.IEEE.ORG

course) to efficiently extract the right digital bits from a host of simultaneous transmissions in which everything looks like noise. (Today all 2G and 3G digital standards use the Viterbi algorithm.) Viterbi and his colleagues also figured out a way to have each handset and tower analyze the quality of its own signals as well as the other conversations being transmitted around it; the tower would then adjust power usage until the handset was transmitting just enough signal to work.

Qualcomm cofounder Jacobs says that it was a long uphill battle for the company to convince the world that its technology, code-division multiple access, or CDMA, was commercially viable, and even the usually patient Viterbi sometimes got frustrated. "A couple of professors at Stanford were being quoted in the press saying that CDMA violated the laws of physics," Jacobs says. "[Viterbi] did get into a bit of an interchange with one of them." For his part Viterbi recalls those years as intense but "very exciting. Things were happening. Hardware was being built. Maybe we were whistling in the dark, but the technology worked."

His daughter, Audrey, theorizes that this battle was perhaps "the most exciting period of his career. My father is always up for a challenge; that's what motivates him." Eventually, Qualcomm's CDMA won out. In 1993, the Telecommunications Industry Association, a trade group that represents about 600 telecom companies, incorporated it into its wireless cellular standard. Today a form of it is used in 3G cellular systems throughout the world.

Viterbi retired from Qualcomm in 2000, with hundreds of millions of dollars in stock. He now had the time and the resources to do just about anything. And what he really wanted, he decided, was to do more to help up-and-coming engineers.

These days, Viterbi's helping to groom the next generation of tech entrepreneurs through his investment firm, the Viterbi


Group, which consists of Viterbi, his daughter Audrey, and an assistant. Not surprisingly, he focuses on communications start-ups, many of which are adapting the fundamental algorithm he discovered to new purposes. He's not trying to control his technology, he explains; it's just the area he knows best. To date, the group has invested in 30 companies and currently has some US \$10 million invested in 10 companies.

One of his earliest investments was in VoiceSignal Technologies, which used the Viterbi algorithm for voice recognition systems in cellphones, including the iPhone. Nuance Communications acquired VoiceSignal in 2007. He was also an early investor in Provigent, a company that produces systems for microwave point-to-point transmission; Flarion Technologies, a company with a 4G phone technology that was eventually acquired by Qualcomm; and TransChip, a camera-on-a-chip manufacturer purchased by Samsung. While Viterbi may be a money man for these companies, his technical advice and industry contacts are invaluable to the start-up teams. Rajiv Laroia, Flarion's founder, says that when his company was still just a struggling New Jersey start-up, Viterbi's stamp of approval conferred instant credibility. "He is a god in the field of communications," says Laroia.

Viterbi has also made generous donations to engineering education at USC (where the school of engineering now bears his name); MIT; the Technion-Israel Institute of Technology, in Haifa; Boston Latin School; and two private high schools.

"The future of this nation is scientific literacy," Viterbi says. "What else do we have to sell? We have innovation and we have a lot of bright kids."

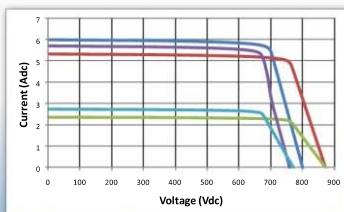
Maybe one of those bright kids will grow up, challenge the established way of doing things, and change the world of technology. Just like Andrew Viterbi. □

**MAGNA-POWER
ELECTRONICS**


2 KW TO 900 KW+ PROGRAMMABLE AC/DC POWER SUPPLIES

**PHOTOVOLTAIC
EMULATION:**


- Automatic V/I profile generation
- Sequence through changing *temperature* and *irradiance*
- Validate MPPT algorithms
- Flexible modeling
- Compatible with our entire product line: 2 kW to 900 kW+
- For more details, visit: www.magna-power.com/solar




User-defined Emulated Solar Characteristics



DC POWER SUPPLIES:

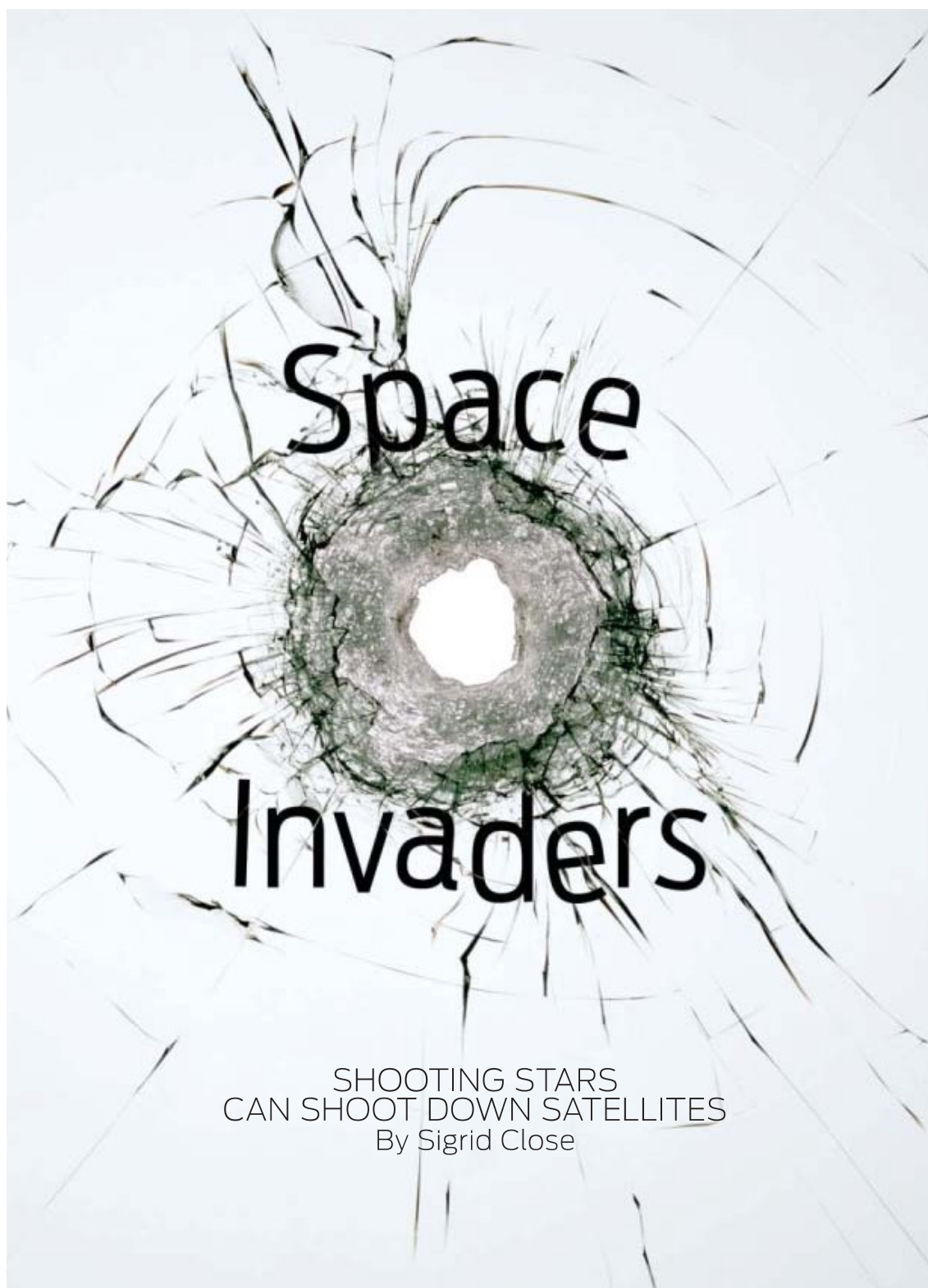
- From 2 kW to 900 kW+
- Fully isolated up to 4000 Vdc
- *Hundreds of models* for every application
- Low output ripple
- US designed and manufactured
- Worldwide distribution
- RS-232, GPIB,  Ethernet programming available

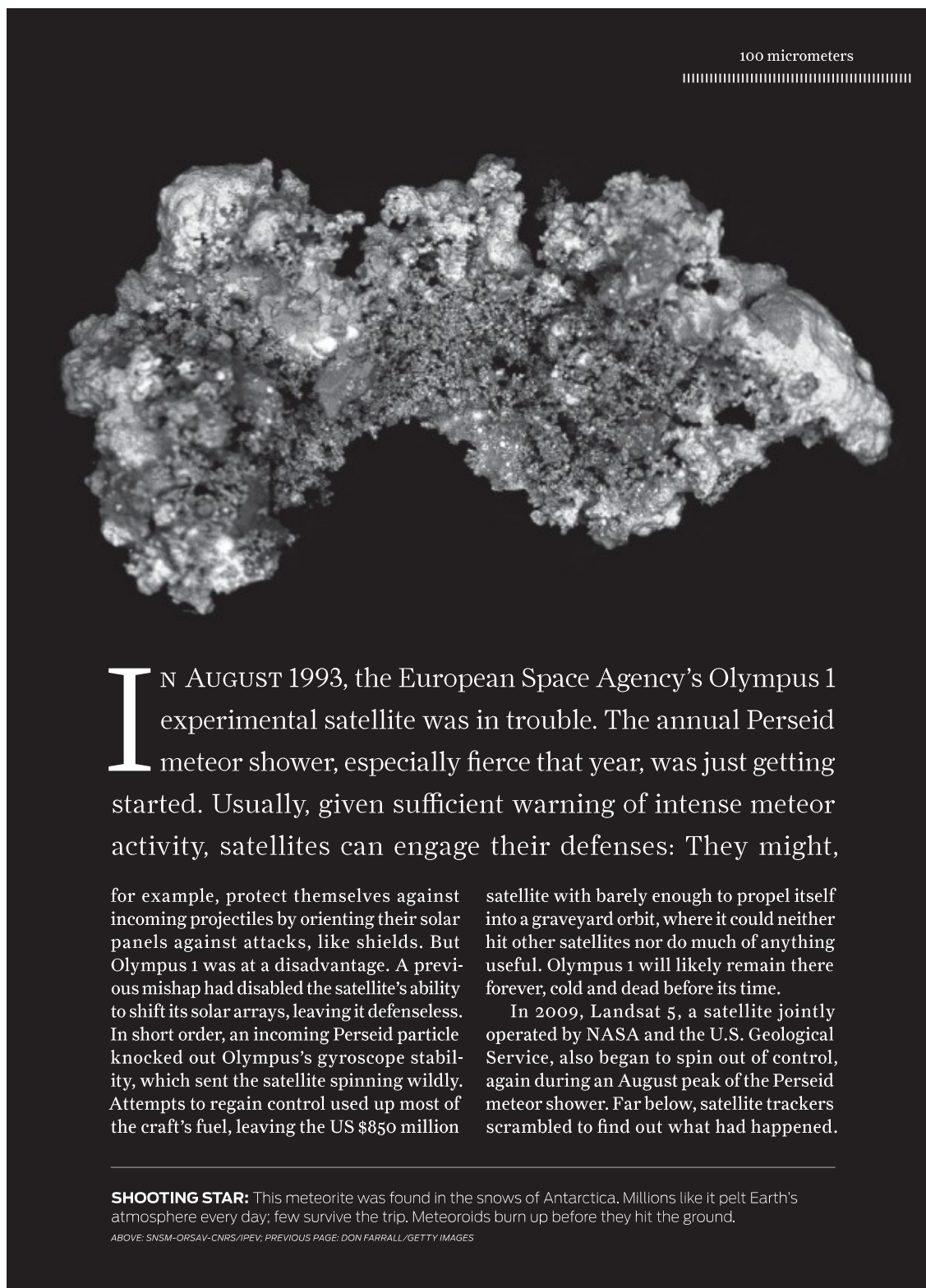


XR Series
Up to 6 kW in a 2U package

**PV
Power
Emulation**

Magna-Power Electronics
39 Royal Road • Flemington, NJ 08822
www.magna-power.com





IN AUGUST 1993, the European Space Agency's Olympus 1 experimental satellite was in trouble. The annual Perseid meteor shower, especially fierce that year, was just getting started. Usually, given sufficient warning of intense meteor activity, satellites can engage their defenses: They might,

for example, protect themselves against incoming projectiles by orienting their solar panels against attacks, like shields. But Olympus 1 was at a disadvantage. A previous mishap had disabled the satellite's ability to shift its solar arrays, leaving it defenseless. In short order, an incoming Perseid particle knocked out Olympus's gyroscope stability, which sent the satellite spinning wildly. Attempts to regain control used up most of the craft's fuel, leaving the US \$850 million

satellite with barely enough to propel itself into a graveyard orbit, where it could neither hit other satellites nor do much of anything useful. Olympus 1 will likely remain there forever, cold and dead before its time.

In 2009, Landsat 5, a satellite jointly operated by NASA and the U.S. Geological Service, also began to spin out of control, again during an August peak of the Perseid meteor shower. Far below, satellite trackers scrambled to find out what had happened.

SHOOTING STAR: This meteorite was found in the snows of Antarctica. Millions like it pelt Earth's atmosphere every day; few survive the trip. Meteoroids burn up before they hit the ground.

ABOVE: SNSM-ORSAY-CNRS/IPEV; PREVIOUS PAGE: DON FARRALL/GETTY IMAGES

Had the satellite been hit by a stray rock from space? Was the culprit a solar storm or an impact with one of the thousands of pieces of orbiting debris shed from other spacecraft or booster rockets? Or was the cause more alarming: an attack from a hostile nation?

Satellite failures such as these have cost the governments of the world billions of dollars. So it might appear strange that there are still huge gaps in our understanding of what can go wrong in space. Space agencies all over the world often struggle to figure out what has happened when some piece of hardware in orbit goes haywire. How can they differentiate between a malfunction caused by a tiny rock hurtling in from interplanetary space and an errant screw left in orbit decades ago? And more important: How can engineers properly protect spacecraft from such projectiles—whatever their provenance?

We live in a time when sending people to Mars is beginning to look like a real possibility, China has announced its plans to set up a moon base by 2030, and soldiers in the field depend on GPS satellites. And that's not to mention regular citizens, who are finding it hard to live without the things satellites provide: long-distance communications, much of their entertainment, and help navigating around town. Therefore, many of us in the space sciences community are renewing our efforts to understand the myriad natural and artificial dangers that spacecraft constantly face.

AFTER 53 YEARS of sending equipment into space, planet Earth has accumulated a thick mantle of space debris. We are able to track about 20 000 objects—although the estimates that account for objects under 10 centimeters in diameter put the total number closer to 600 000. When a satellite hits an object in this belt, the collision may cause the satellite to splinter into many fragments, which then add to the accumulating debris. And because the satellite has been destroyed or crippled, it becomes necessary to send a replacement, increasing the potential for more debris later on.

Satellite designers can fashion shields that guard reasonably well against impacts with small pieces of space debris. And satellite operators can usually track and avoid the larger chunks, although sometimes that process doesn't go so smoothly. Within the past year, both the space shuttle *Discovery* and the International Space Station had to take rapid evasive action to dodge one especially treacherous object. And in February 2009, Russia's defunct Cosmos 2251 satellite collided with an Iridium Communications satellite, unlikely to be the last accident of its kind. Such incidents have prompted spaceflight operators to coin the term "space situational awareness." It's an acknowledgment that you need to see and understand what's going on in space. The growing focus on space situational awareness comes partly from the sheer number of satellites now in orbit and partly from the vulnerabilities inherent in the crowded, busy lanes of near-Earth space.

Adding to these concerns, a 2007 Chinese antisatellite test showed that even this newcomer to the space race is capable of destroying a target in low orbit using a ground-based missile. Another worry is the growing popularity of small satellites, often dubbed microsats or nanosats—names that don't reflect their real size—which typically measure a few centimeters. Because these objects are hard to track and maintain in their proper orbits, they pose headaches for the people whose job it is to catalog everything circling Earth.

In the United States, the Air Force and NASA bankroll most of the fundamental research on reducing the threat of collision in space. Both organizations have obvious interests in being able to travel in space or to place useful objects into orbit.

IHAVE WORKED with Bill Cooke at NASA's Marshall Space Flight Center and with researchers at the Air Force's Office of Scientific Research throughout my career, first while I conducted studies at MIT Lincoln Laboratory and later at Los Alamos National Laboratory, where my team searched for ground-based electromagnetic pulses using LANL's radio frequency sensors, and now as an assistant professor at Stanford University. All the while, my job has been to learn about the dangers to spacecraft. I try to understand the many complications that can make a satellite mission go awry, including run-ins with orbital debris, lightning, and solar flares. My primary focus, however, is meteoroids, which are solid extraterrestrial bodies smaller than a boulder but larger than a dust grain.

Satellite engineers have done much to mitigate space hazards, creating among other things the Whipple bumper, a kind of hypervelocity impact shield that's designed to be as light as possible. You might guess that the biggest threats are the relatively bulky objects that make up space debris, which in low Earth orbit are bigger and more numerous than typical meteoroids. But meteoroids make up for their small size with high velocities. They tend to travel in the neighborhood of 12 to 72 kilometers per second when entering Earth's atmosphere and can penetrate shielding more easily than a comparatively sluggish piece of space debris can. A Whipple shield can protect a satellite from space junk hitting at speeds up to 18 km/s, but it's no match for faster meteoroids.


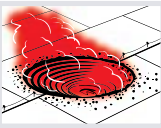


Given the danger they represent, it's downright shocking how much remains unknown about meteoroids. Ironically enough, what's best understood about them is how they act once they plunge into the atmosphere and become visible, when they are no longer a threat to spacecraft.

EVERY DAY, MORE THAN 100 BILLION meteoroids larger than one microgram enter Earth's atmosphere, traveling at more than 11 km/s. By the time these have made their way through the ionosphere (a plasma that extends between 70 and 1000 km above Earth's surface, its height depending on factors like solar cycle and time of day), the vast majority of meteoroids have been scoured away to nothing by the friction of the thickening atmosphere around them. As the mass is removed, it fans out behind the nucleus, forming a long glowing tail of plasma. These quick-moving flashes of light are commonly called shooting stars, although the correct term is *meteor*.

Meteor showers take place when Earth passes through the orbit of cometary particles. A shower is named after the constellation from which it appears to come. The Perseids seem to emanate from the constellation Perseus, the Leonids from the constellation Leo. Satellite operators are very familiar with these periodic barrages, and for the most part, they know how to protect their valuable space-borne assets.

Although it would be impossible to shield a satellite from a good-size cobble speeding along at many kilometers per second, such things are too rare to cause significant concern. The threats are the really small meteoroids—under 0.05 millimeter in diameter—which exist in much greater numbers. These interplanetary flyspecks aren't big enough to make flashes in the atmosphere visible to the naked eye;

Sizing Up the Space-Debris Hazard

	Event	Diameter (meters)	Impact energy (megatons)	Frequency
	Shooting star	0.00006	5×10^{-16}	1 second
	Brilliant fireball	0.1	0.01	1 year
	Aerial burst with modest damage	25	1	200 years
	Local (10-km) devastation	50	10	2000 years
	Regional-scale devastation	140	300	30 000 years
	Continental-scale devastation	300	2000	100 000 years
	Possible global catastrophe	1000	100 000	700 000 years
	Global extinction	10 000	100 million	100 million years

they're so tiny that spacecraft designers have not generally considered them much of a threat. But for satellites, they can be lethal. The damage these tiny grains inflict comes in part directly from the holes they make. Although they have little mass, they can travel extraordinarily fast. So even infinitesimal meteoroids can pack quite a punch.

Exactly how fast these meteoroids travel has been a bit of a mystery. For a long time, scientists simply assumed that the dominant population of submilligram meteoroids travel at around 20 km/s. But recent data gathered at high-power, large-aperture radar facilities (such as the Arecibo Observatory, EISCAT Scientific Association, Jicamarca Radio Observatory, and ARPA Long-Range Tracking and Instrumentation Radar, or ALTAIR) suggest the typical speed for meteoroids smaller than 50 micrometers is closer to 60 km/s. That's a pretty large correction.

We've also learned that the mass of a meteoroid, as well as its composition, depends in part on where it comes from. The prevailing wisdom has been that none of this material hails from beyond our own solar system, but I and others

have recently done work that challenges this long-held belief. Although our conclusion courts controversy, we think that at least 4 percent may come from interstellar space, from the exploding stars that create pulsars and from other exotic locales, like the dust-enshrouded star Beta Pictoris, 63.4 light-years from our solar system. These are impressively distant origins for a little nub of matter that could easily blow a hole in a billion-dollar satellite. Interstellar meteoroids are faster than the fastest meteoroids from inside the solar system, entering Earth's atmosphere at speeds far greater than even the 72.8 km/s that most scientists currently define as the limit for a meteoroid originating outside our solar system.

BY NOW YOU'RE PROBABLY wondering why satellite designers can't just build better shielding and be done with it. Better physical shielding would help, of course. But it wouldn't do anything to protect spacecraft from a more subtle problem: the electrostatic discharges and electromagnetic pulses that accompany many impacts. The physics of these electrical effects is complicated. Even the notion that electro-

ILLUSTRATIONS: MCKIBELLO

Kill Your Satellite

METEOROIDS HAVE TAKEN out more than a few spacecraft. In addition to Olympus and Landsat 5, other possible victims were the Small Expendable Deployer System (March 1994) and the Miniature Sensor Technology Integration (also March 1994).

Two types of electrical effects are associated with an impacting meteoroid: electrostatic discharges (ESD) and electromagnetic pulses (EMPs). An ESD—you'll know it as the garden-variety spark—occurs on a satellite when accumulated electric charge is suddenly discharged. It's the same principle by which you experience a mild electrical shock



when you touch something after you've been walking across a carpeted floor. The satellite builds up charge simply by traveling in its orbit. Then, when the meteoroid smacks into the charged-up satellite, it acts as a flint to generate the spark.

An ESD is a serious problem for the many integrated circuits on board a satellite. These devices are made from semiconductors like silicon and gallium arsenide, materials that can sustain permanent damage from high voltages. The result of an impact could be a localized ESD, or a more serious problem such as a discharge of the whole satellite, depending on the material of the spacecraft. Some antistatic devices can help prevent static buildup, but no one knows how much buildup

magnetic radiation from meteoroid impacts—tiny versions of the electromagnetic pulses given off by nuclear bombs—could be sufficiently energetic to destroy circuitry is controversial. But some NASA engineers suspect that exactly such an event led to the untimely demise of Landsat 5's gyros last year. Those who disagree think the culprit was the direct physical damage. But one thing is clear: Space scientists need to learn a lot more about meteoroids and the dangers they pose.

One way to study meteoroids is to observe how they shift the orientation of the satellites they strike. Scientists can use this simple method to detect meteoroid collisions, because the angular momentum of an object in orbit doesn't change without a reason. If a satellite's velocity shifts (and if other variables, like gravity, light pressure, and atmospheric drag are ruled out), the only logical conclusion is that something has hit it. It's easy to apply this technique with ALTAIR, because this radar can determine velocities with extreme precision. But even so, it's impossible to tell the difference between a satellite's collision with a large piece of relatively slow-moving space debris and a collision with a small but speedy meteoroid. Both could impart exactly the same change in momentum.

A better way to differentiate meteoroids from space debris—and to distinguish among the different kinds of meteoroids—is sorely needed. A technique that could do that would also help solve another, related problem: not being able to tell the difference between naturally occurring phenomena in our own ionosphere and man-made artifacts like launch vehicles or missiles. A good approach, I and many other space scientists believe, is

to try to gauge the size and speed of all this space flotsam and jetsam by looking carefully at the plasma trails these objects create when they hit the atmosphere. With ALTAIR, that's a fairly straightforward exercise because the plasma reflects radio waves so well. Of course, you need to model the object's motions as it burns up to calculate how fast it was going in the first place. But if you do this right, you can also figure out its mass, density, and radius. This method is certainly better than just measuring the momentum transferred to satellites when they are struck. But the best strategy of all would be to observe

Deep Impact



METEOROID DAMAGE: Pictured are the results of impacts on the Atlantis space shuttle payload bay door radiator, an insulation blanket on a module of the International Space Station, and the Hubble Space Telescope solar array.

BETWEEN 1992 and 2002, the space shuttles were assessed for meteoroid and debris damage 50 times. (Although the craft are assessed after each mission, there are more in-depth examinations if something is found to raise interest; think of it as the difference between regular Transportation Security Administration screenings and being pulled over for inspection.) After 62 missions, NASA inspectors found that the shuttle windows alone had been struck 1578 times and that 98 of these incidents led to damage that required repairs. Spectroscopic analysis of these

98 sites—in which inspectors looked at trace materials left on the collision site—revealed that orbital debris had caused 41 impacts. Eighteen were attributable to meteoroids; the causes of the other collisions are either unknown or left no samples in the impact craters. In low Earth orbit, where the shuttle operates, we expect human-made space debris to exceed the meteoroid population. But the space around satellites that lie farther out, in geosynchronous orbit, teems with equally large populations of space debris and meteoroids. By some estimates, in fact, meteoroids are more prevalent. —S.C.

meteoroid impacts up close from a vantage point in space.

To that end, I am collaborating with Andrew Kalman at Stanford to develop a satellite called MEDUSSA, which stands for Meteoroid and Energetics Detection for Understanding Space Situational Awareness. If it is built and flown as I hope, the MEDUSSA satellite will be able to study exactly what takes place when micrometeoroids and energetic particles slam into it.

With this information, engineers should be better able to design satellites to resist impact damage. Of course, there will still be run-ins with meteoroids and orbital debris that destroy spacecraft in mysterious ways, just as there are disasters with aircraft that defy explanation. But MEDUSSA would certainly help. And one day down the road, I anticipate that every satellite that gets lofted will contain a stand-alone unit to sense impacts and report their effects—even if they are extreme enough to disable the rest of the satellite. Think of it as a satellite “black box.” Maybe then, we'll get a good picture of just all the unexpected things that can go wrong in space. □

This article would not have been possible without the invaluable contributions of Stan Green.

they need to prevent because these effects have not been thoroughly studied.

Unlike ESDs, electromagnetic pulses (EMPs) occur when a colliding meteoroid vaporizes on impact and forms a plasma. In a collision, both the meteoroid and a fraction of the satellite evaporate and ionize, instantaneously forming a cloud of plasma. That plasma can then affect equipment such as the RF antenna on board.

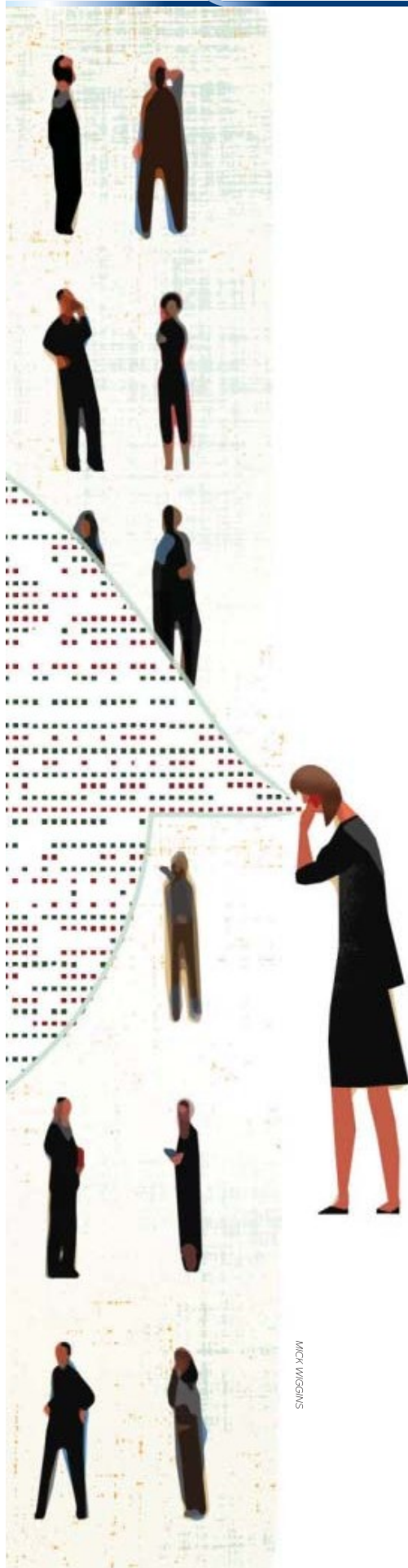
The theory that meteoroid-induced EMPs cause catastrophic satellite failures is controversial and, at this writing, unconfirmed. However, researchers have been studying the plasma production associated with hypervelocity impacts for over 30 years. Enough unsolved satellite anomalies have taken place to make the case for EMPs as a damage mechanism. —S.C.

VICE OVER IP

A growing cadre of criminals is hiding secret messages in voice data

By JÓZEF LUBACZ,
WOJCIECH
MAZURCZYK
& KRZYSZTOF
SZCZYPIORSKI





MICK WIGGINS

7:00 P.M., SHANGHAI

An employee of an electronic equipment factory uploads a music file to an online file-sharing site. Hidden in the MP3 file (Michael Jackson's album *Thriller*) are schematics of a new mobile phone that will carry the brand of a large American company. Once the employee's Taiwanese collaborators download the file, they start manufacturing counterfeit mobile phones essentially identical to the original—even before the American company can get its version into stores.

3:30 P.M., SOMEWHERE IN AFGHANISTAN

A terrorist hunted by the U.S. Federal Bureau of Investigation posts an excerpt from the motion picture *High School Musical Three: Senior Year* on Facebook. Inside are hidden instructions for a bomb attack on a commuter rail line in southern Europe. Later that day, terrorists based in Athens follow the instructions to plan a rush hour attack that kills hundreds of people.

4:00 A.M., MALIBU, CALIF.

A very famous actor (VFA) has a brief conversation with a well-known director (WKD) over Skype, an application that lets them make free voice calls over the Internet. They discuss the medical problems of VFA's cat in great detail. When the conversation is over, WKD's computer has a sleazy new addition—in a folder on his desktop, there is a picture of a nude teenager, along with her mobile number and the date and time at which WKD will meet her at VFA's pool party for a photo session.



WHAT ALL these scenarios have in common is an information-smuggling technique called steganography—the communication of secret messages inside a perfectly innocent carrier. Think of steganography as meta-encryption: While encryption protects messages from being read by unauthorized parties, steganography lets the sender conceal the fact that he has even sent a message. After the 11 September attacks in 2001, rumors flew that they had been carried out with some help from steganography. A 2001 *New York Times* article described fake eBay listings in which routinely altered pictures of a sewing machine contained malevolent cargo. The link to 9/11 was never proved or disproved, but after those reports, the interest in steganographic techniques and their detection greatly increased.

Steganography use is on the rise, and not just among criminals, hackers, child pornographers, and terrorists. Persecuted citizens and dissidents under authoritarian regimes use it to evade government censorship, and journalists can use it to conceal sources. Investigators even use it on occasion to bait and trap people involved in industrial espionage: In the 1980s, to trace press leaks of cabinet documents, British Prime Minister Margaret Thatcher had government word processors altered to encode a specific user identity in the spaces between words. When leaked material was recovered, the identity of the leaker could be established by analyzing the pattern of those spaces.

Steganography is evolving alongside technology. A few years ago the cutting edge in steganographic tools involved hiding messages inside digital images or sound files, known as carriers, like that *Thriller* MP3. The technique quickly evolved to include video files, which are relatively large and can therefore conceal longer messages.

Now steganography has entered a new era, with stupendously greater potential for mischief. With the latest techniques, the limitations on the length of the message have basically been removed. Consider our example involving the use of Skype. Whereas the first two examples each required a carrier—an MP3 song and a video—there was no such requirement for the transmission of that nude photo. The data were secreted among the bits of a digital Voice over Internet Protocol conversation. In this new era of steganography, the mule that coconspirators are using is not the carrier itself but the communication protocols that govern the carrier's path through the Internet. Here's the advantage: The longer the communicators talk, the longer the secret message (or more detailed the secret image) they can send.

CARRIER EVOLUTION

Steganography has been used for at least 2500 years to disguise secret messages. In its earliest forms, the carriers were physical, but as technology evolved, so did carriers.

494 B.C. HEAD TATTOO



Histiaeus tattoos a secret message onto a slave's shaved head, waits for the hair to regrow, and sends the slave to the intended recipient, who shaves off the hair to read the message.

480 B.C. BEESWAX

Demaratus writes a secret message on a wooden tablet to warn the Greeks of Persian attack, and then covers it with many coats of wax.

1558 EGGS

Italian scientist Giambattista della Porta discovers how to hide a message inside a hard-boiled egg: Write on the shell using an ink made from a mixture of alum and vinegar. The solution leaves no trace on the surface,

Most strikingly, the concealment occurs within data whose inherent ephemerality makes the hidden payload nearly impossible to detect, let alone thwart.

We call this new technique network steganography. In our research at the Network Security Group at Warsaw University of Technology, we are studying the ever-evolving spectrum of carrier technologies, the increasing difficulty of detection as more sophisticated carriers leave fewer traces, and the implications of both for law enforcement and homeland security. Our work at Warsaw is literally self-defeating: We figure out the most advanced ways of doing network steganography and then design methods to detect them.

NETWORK STEGANOGRAPHY is a modern version of an old idea. You could argue that steganography helped spark the first major conflict between Greece and the Persian Empire. A classic use of steganography took place in 494 B.C., when Histiaeus, the ruler of Miletus, tried to instigate an Ionian revolt against the Persians. He shaved his favorite slave's head, tattooed it with a message, and waited for the slave's hair to grow back and obscure the tattoo. Then he sent the slave to his destination, where the intended recipient shaved the slave's head and read the message. The ensuing Ionian revolution lasted for half a century. In the 19th and 20th centuries, rapidly evolving warfare and espionage brought many innovations in steganography: Invisible ink, microdots, and Thatcher's word-processor trick are only a few among many.

With today's technology, information can be smuggled in essentially any type of digital file, including JPEGs or bitmaps, MP3s or WAV files, and MPEG movies. More than a hundred such steganographic applications are freely available on the Internet. Many of these programs are slick packages whose use requires no significant technical skills whatsoever. Typically, one mouse click selects the carrier, a second selects the secret information to be sent, and a third sends the message and its secret cargo. All the recipient needs is the same program the sender used; it typically extracts the hidden information within seconds.

A SINGLE 6-MINUTE MP3 OCCUPIES 30 MB,
ENOUGH TO CONCEAL EVERY PLAY SHAKESPEARE EVER WROTE

Any binary file can be concealed—for instance, pictures in unusual formats, software (a nasty virus, say), or blueprints. The favored carrier files are the most common ones, like JPEGs or MP3s. This emphasis on popular file formats increases the anonymity of the entire transaction, because these file types are so commonplace that they don't stick out.

The one limitation that steganographers have traditionally faced is file size. The rule of thumb is that you can use 10 percent of a carrier file's size to smuggle data. For an ambitious steganographer, that could be a problem: Imagine an electronic equipment factory employee trying to explain to the IT department why he has to send his mother a 100-megabyte picture of the family dog. For that reason, steganographers soon turned to audio and video files. A single 6-minute song, in the MP3 compression format, occupies 30 MB; it's enough to conceal every play Shakespeare ever wrote.

And yet, even with these precautions, conventional steganography still has an Achilles' heel: It leaves a trail. Pictures and other e-mail attachments stored on a company's outgoing e-mail servers retain the offending document. Anything sent has to bounce through some kind of relay and can therefore be captured, in theory.

Steganography poses serious threats to network security mainly by enabling confidential information leakage. The new crop of programs leaves almost no trail. Because they do not hide information inside digital files, instead using the protocol itself, detecting their existence is nearly impossible.

ALL THE new methods manipulate the Internet Protocol (IP), which is a fundamental part of any communication, voice or text based, that takes place on the Internet. The IP specifies how information travels through a network. Like postal service address standards, IP is mainly in charge of making sure that sender and destination addresses are valid, that parcels reach their destinations, and that those parcels conform to certain guidelines. (You can't send e-mail to an Internet address that does not use a 32-bit or 128-bit number, for example.)

but the message is retrieved by removing the shell and reading the egg.

1800s
NEWSPAPER
CODE

During the Victorian era, lovers send secret letters by punching holes above certain letters. When the marked letters are combined, the message can be read.

1915
INVISIBLE INK

During World War I, entertainer and German spy Courtney de Rysbach performs in shows all over Britain as a cover for gathering information. Using invisible ink, Rysbach encodes secret messages by writing them in invisible ink on sheets of music.

1941
MICRODOTS

During World War II, German agents photographically shrink a page of text down to a 1-millimeter dot. The microdot is then hidden on top of a period in an otherwise unremarkable letter.

All traffic, be it e-mail or streaming video, travels via a method called packet switching, which parcels out digital data into small chunks, or packets, and sends them over a network shared by countless users. IP also contains the standards for packaging those packets.

Let's say you're sending an e-mail. After you hit the Send button, the packets travel easily through the network, from router to router, to the recipient's in-box. Once these packets reach the recipient, they are reconstituted into the full e-mail.

The important thing is that the packets don't need to reach their destination in any particular order. IP is a "connectionless protocol," which means that one node is free to send packets to another without setting up a prior connection, or circuit. This is a departure from previous methods, such as making a phone call in a public switched telephone network, which first requires synchronization between the two communicating nodes to set up a dedicated and exclusive circuit. Within reason, it doesn't matter when packets arrive or whether they arrive in order.

As you can imagine, this method works better for order-insensitive data like e-mail and static Web pages than it does for voice and video data. Whereas the quality of an e-mail message is immune to traffic obstructions, a network delay of even 20 milliseconds can very much degrade a second or two of video.

To cope with this challenge, network specialists came up with the Voice over Internet Protocol (VoIP). It governs the way voice data is broken up for transmission the same way IP manages messages that are less time sensitive. VoIP enables data packets representing a voice call to be split up and routed over the Internet.

The connection of a VoIP call consists of two phases: the signaling phase, followed by the voice-transport phase. The first phase establishes how the call will be encoded between the sending and receiving computers. During the second phase, data are sent in both directions in streams of packets. Each packet, which covers about 20 milliseconds of conversation, usually contains 20 to 160 bytes of voice data. The connection typically conveys between 20 and 50 such packets per second.

Telephone calls must occur in real time, and significant data delays would make for an awkward conversation. So to ferry a telephone call over the Internet, which was not originally intended for voice communications, VoIP makes use of two more communications protocols, which had to be layered on top of IP: The Real-Time Transport Protocol (RTP) and the User Datagram Protocol (UDP). The RTP gets time-sensitive video and audio data to its destination fast and so has been heavily adopted in much of streaming media, such as telephony, video teleconference applications, and Web-based push-to-talk features. To do that, it relies in turn on the UDP.

Because voice traffic is so time critical, UDP does not bother to check whether the data are reliable, intact, or even in order. So in a VoIP call, packets are sometimes stuck in out

ALL THREE STEGANOGRAPHIC IDEAS WE'VE OUTLINED HERE ARE SO SIMPLE, WE'RE CERTAIN THAT REAL-LIFE APPLICATIONS ARE ALREADY OUT THERE

of sequence. But that's not a big deal because the occasional misplaced packet won't significantly affect the quality of the phone call. The upshot of UDP is that the protocol opens a direct connection between computers with no mediation, harking back to the era of circuit switching: Applications can send data packets to other computers on a connection without previously setting up any special transmission channels or data paths. That means it's completely private.

Compared to old-fashioned telephony, IP is unreliable. That unreliability may result in several classes of error, including data corruption and lost data packets. Steganography exploits those errors.

Because these secret data packets, or "steganograms," are interspersed among many IP packets and don't linger anywhere except in the recipient's computer, there is no easy way for an investigator—who could download a suspect image or analyze an audio file at his convenience—to detect them.

TO BETTER UNDERSTAND what security officials will soon have to deal with, we designed and developed three flavors of network steganography, all of which manipulate IP. The three methods we developed are Lost Audio Packet Steganography, or LACK; Hidden Communication System for Corrupted Networks (HICCUPS); and Protocol Steganography for VoIP application. As their names imply, these techniques exploit lost packets, corrupted packets, and hidden or unused data fields in the VoIP transmission protocol. LACK hides information in packet delays, HICCUPS disguises information as natural "distortion" or noise, and Protocol Steganography hides information in unused data fields.

In regular VoIP telephony, excessively delayed packets containing voice samples are considered useless by the receiver and thus discarded. LACK exploits this mechanism to transmit hidden data. Some of the sender's packets are intentionally delayed, and the steganograms are stowed away inside those delayed packets. To any node that is not "in the know"—that is, a nearby computer that does not have the steganography program installed—they appear useless and are ignored. But if the receiver has the proper software to understand the steganography, it will not discard the excessively delayed packets. It will know that these contain the hidden data [see diagram, "Hidden in the Network"].

The transmission capacity for this scheme depends on the system used to encode the voice and on the quality of the network—specifically, how well it handles packet loss and delays. Using a standard 32-bit-per-second codec, and accounting for a 3 percent packet loss introduced by the network and a 0.5 percent packet loss introduced by LACK itself, a smuggler could transmit about 160 bits per second. At that rate you might be able to transmit a medium-size, 13-kilobyte image or a 2000-word text file during a typical 9- to 13-minute VoIP conversation.

LACK's main selling points are that it is simple to use and hard to detect. The only way it could be detected is if the user tried to hide too many secret packets. In that case, the

1980s

WATERMARKING



In the 1980s, to trace press leaks of cabinet documents, British Prime Minister Margaret Thatcher has government word processors altered to encode a specific user identity in the spaces between words.

1990s

DIGITAL STEGANOGRAPHY

Researchers develop methods to secretly embed a signature in digital pictures and audio, exploiting the human visual system's varying sensitivity to contrast.

2003

STREAMING VIDEO

Video steganography is similar to image steganography, but more information may be transported in a stream of images.

2007

NETWORK STEGANOGRAPHY

New methods focus on using free or unused fields in a protocol's headers.

number of intentionally delayed packets—and therefore the introduced delay—would create a suspiciously abnormal voice connection that might attract the attention of any security officials monitoring the line. If the call was completed before those officials could intercept the packets, however, there would be nothing they could do to try to uncover and assemble the steganograms.

Where LACK relies on lost packets to smuggle steganograms, HICCUPS takes advantage of corrupted packets. HICCUPS is fast. Let's say you have an IEEE 802.11g network with a transmission capacity of 54 megabits per second, with 10 terminals and a 5 percent rate of corrupted frames. Over such a network, you could send hidden data at a rate higher than 200 kilobits per second. That's almost as fast as the ISDN lines that were all the rage in the 1990s.

HICCUPS works on wireless local area networks, such as plain old coffee shop Wi-Fi. In such a wireless environment, data are transmitted by a method called broadcasting, which shuttles data in groups called frames. Like many courier services, broadcasting doesn't concern itself with the contents of the data or whether the data contain errors. When a wireless network detects an error in a frame, the computer simply drops that corrupted frame. The responsibility for detecting dropped frames (and retransmitting them if necessary) is left to the origin and destination terminals.

So in a wireless local-area network, all the user terminals (laptops, for the most part) must have a way of differentiating good packets from corrupted ones. This error-checking mechanism is called the checksum, a kind of signature against which the integrity of the packets can be confirmed. The checksum is a numerical value assigned to a data packet based on the number of bits in that packet. A checksum program uses that value to authenticate that the data hasn't been corrupted.

When the receiver's computer gets a packet, it checks for errors using that packet's checksum. Normally, if the checksum is wrong, the computer discards that packet. But if a terminal has

AP PHOTO



HIDDEN IN THE NETWORK

the right steganography program installed, it won't discard these intentionally wrong checksums—instead, it will know that these are precisely the data packets to scan for steganograms.

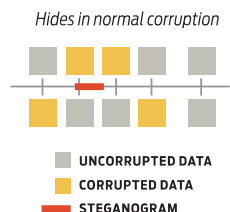
HICCUPS is more difficult to pull off than LACK. That's because this method requires a wireless card that can control frame checksums (good luck finding one of those at RadioShack). Network cards create checksums at the hardware level. We have applied for a patent in Poland for a HICCUPS-enabled card that can control checksums, but so far we haven't built our own card. Detecting HICCUPS wouldn't be easy. You'd need some way of observing the number of frames with incorrect checksums. If the number of those frames is statistically anomalous, then you might suspect the transmission of hidden information. Another way of detecting HICCUPS would analyze the content of those dropped—and therefore retransmitted—frames in order to detect the differences between the dropped and retransmitted frames. Major differences in these frames would constitute an obvious clue to nefarious goings-on.

Any of these detection methods, of course, would require not only that an investigator be aware that a transmission was about to take place but also that he be equipped with the right equipment, ready to monitor the conversation and intercept bits. Such a situation would be unlikely, to put it mildly.

The third method, Protocol Steganography, is a common name for a group of methods that use another aspect of IP: packet header fields. These fields are like sophisticated address labels that identify the contents of data packets to the recipient. Steganograms can be hidden inside unused, optional, or partial fields, because any data in these fields can be replaced without affecting the connection. Some of the more ham-fisted steganography techniques simply replace the content of the unused or optional fields with steganograms. But that would be relatively easy to detect and even jam.

So, to evade detection by simple analysis, the more sophisticated variant of Protocol Steganography uses fields in which the content changes frequently. For example, some of the more esoteric VoIP fields carry security data for authentication purposes. That little authentication subfield changes frequently during the course of a normal call. A steganogram smuggled inside one of its many randomly changing packets would be extremely hard to detect. Of course, there is a trade-off: The user would also sacrifice security, meaning that his or her conversation could be intercepted more easily.

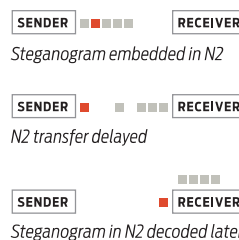
Minimizing the threat of evolving steganography methods

HICCUPS
(CORRUPTED PACKETS)

Highest information density
HICCUPS [red] hides in the "noise" of natural distortion [orange] in an otherwise normal VoIP telephone call [gray].

Difficult to use Because this method requires hardware that can generate wrong checksums, it is difficult to use.

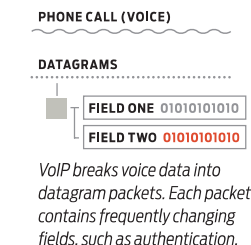
200 kilobits per second are transmitted during a typical 9–13 minute VoIP call.

LACK
(LOST AUDIO PACKETS)

Lowest information density
Excessively delayed packets are dropped by the receiver. LACK delays packets on purpose, encodes the hidden data, and decodes the steganograms when they arrive.

Hardest to detect Used carefully, LACK delays only a small percentage of packets.

160 bits per second are transmitted during a typical call.

PROTOCOL
STEGANOGRAPHY
(HIDDEN FIELDS)

Easiest to use Each bit (phone-call data) contains data fields. Some fields contain frequently changing data, which can be wholly or partially replaced with a steganogram.

Hard to detect By replacing the authentication field, the user sacrifices security.

1–300 bits per second are transmitted during a typical call.

requires an in-depth understanding of how network protocols function and how they can be exploited to hide data. The problem is, however, the complexity of today's network protocols. All three steganographic ideas we've outlined here are so simple, we're certain that real-life applications are sure to come, if they aren't already out there. In fact, much more sophisticated methods will appear as Internet communication evolves from VoIP to other real-time media communications, such as video chat and conferencing.

THE ANONYMITY OF STEGANOGRAPHY might be good for privacy, but it also multiplies the threats to individuals, societies, and states. The trade-off between the benefits and threats involves many complex ethical, legal, and technological issues. We'll leave them for other thinkers and other articles.

What we're trying to do is understand what kind of potential contemporary communication networks have for enabling steganography, and in effect, create new techniques so that we can figure out how to thwart them. Some readers may object to our detailed descriptions of how these methods can be harnessed. But we would counter that unless someone shows how easy all this is, researchers won't understand the urgency and be inspired to develop protective measures. Not only can VoIP steganography be implemented in telephony tools that require a laptop or PC (like Skype), it can also be used in hard phones, such as the Android VoIP-enabled mobile phones that are starting to proliferate. Steganography on a phone is more difficult, because it requires access to the device's operating system, but no one should doubt that committed individuals will have no trouble rising to the challenge. As George Orwell once wrote, "On the whole human beings want to be good, but not too good, and not quite all the time." □

Lite, Brite Displays

Kindle, iPad, Droid—these compact mobile devices are essentially all display. But the screens aren't all we'd like them to be. Yet.

By JASON HEIKENFELD

It's 2020, and it's sunny outside. In fact, it's so bright in your kitchen that you have to squint to see your grapefruit. You flip on your e-reader and the most recent e-issue of *IEEE Spectrum* pops up on-screen, the colors and text sharp and brilliant in the sunlight. There's e-mail to answer, but you want to make the early commuter bus, so you roll up your e-reader and stuff it in your jacket pocket.

On the bus, you switch the device to physically rigid mode and half the screen becomes a large keyboard. You bang out a few messages, then watch a short video. All the while the unit is charging its battery through a built-in organic solar cell.

That's my vision of the future of periodical literature—or rather, the future of periodical delivery. It combines the orderly, portable, full-color format of today's print publications with the flexibility, timeliness, and multimedia capabilities of online magazines. And the only component still lacking is a screen that's easy on the eyes in all sorts of lighting conditions, displays full-motion and full-color images, is rollable and durable, and uses precious little power.

Like the jet pack, it always seems to be a decade away. So why should you believe me now when I tell you that the do-all e-reader will be available in a decade? Read on.

NO FEWER THAN HALF A DOZEN different technologies are emerging from laboratories to compete to be the e-reader screen of the future. The stakes are high: Research firm DisplaySearch estimates that the market will near US \$10 billion by 2018, powered by a compound annual growth rate of 41 percent.

To understand the technical challenges, first consider where we are today. Today's electronic readers, such as the Amazon Kindle and the Sony Reader, meet two of my criteria for the ideal e-reader: They're easy to read in bright light and use minimal power. These monochrome displays, sometimes called electronic paper or e-paper, use a kind of electrophoretic technology developed by E Ink Corp., a company in Cambridge, Mass., that was spun out of the MIT Media Lab in 1997. An electrophoretic pixel comprises numerous tiny capsules that contain a mixture of oppositely charged pigment particles, typically carbon for black and titanium dioxide for white. A voltage attracts or repels the pigment particles within the capsules from the screen, depending on whether a white or a black pixel is needed at that spot. Like mixing paints, with the right voltage control the system can also

leave the particles in a partially mixed, or grayscale, state. It doesn't need much power, because the pigments simply reflect—or don't reflect—the ambient light, and they don't need any power to maintain their most recent state. An electrophoretic display takes 200 milliseconds to switch images. So if the image on the display changes every 60 seconds, in 1000 hours of continued use the display would effectively draw power for only about 3 hours.

E Ink has spent over a decade getting to this point and is still refining the basic technology. But already these displays are really simple to produce. Manufacturers purchase ready-made film containing the pigment-filled capsules and simply laminate it to an underlying panel that carries the drive circuitry. The first generation of E Ink displays used silicon transistors and glass panels; the second, due this year, will use organic transistors and plastic panels. This second generation includes Polymer Vision's RADIUS and Plastic Logic's Que; the RADIUS is literally paper thin, and it can be rolled and unrolled tens of thousands of times.

Now for the downside. Electrophoretic technology has limited potential for displaying full-color images. That's because it hasn't really solved the brightness challenge. Imagine that you're going to paint a wall white that's now a very dark brown. You'll need at least three coats of white paint to cover that brown. Electrophoretic pixels have a similar problem, because the black particles are never fully hidden by the white ones. So the white reflectance is only about 40 percent, compared to 80 percent for a sheet of paper.

If you try to get around this problem by using more particles, you run into problems with switching speed. Electrophoretic pixels already switch slowly because the layer of electrophoretic ink is relatively thick, about 40 micrometers, and the voltage applied to the pixel must be spread across the entire thickness. The level of liquid crystal material in LCDs is only a few micrometers thick, and that's one reason they're so much faster. Electrophoretic technology also can't do video; the switching speed is just too slow.

OPENING IMAGE: SEAN MCCABE

Screen Play

At least six technologies are in contention to be the display of the future



ELECTROPHORETIC PIXEL

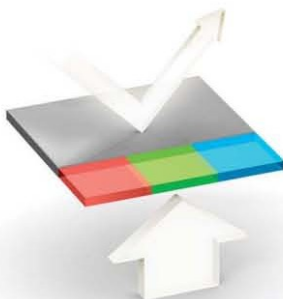
Black and white pigment particles with opposite charges migrate inside capsules, depending on the applied voltage.

PRO Is simple to produce.

CON Slow switching limits video capability; full color is dim.

WHO E Ink

STATUS Available now



3QI MULTIMODE

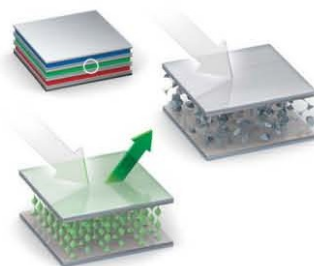
This LCD variant both reflects and transmits light.

PRO Consumes minimal power and is visible indoors and out.

CON Brightness and color saturation are both compromised.

WHO Pixel Qi

STATUS Expected to ship this year



CHOLESTERIC LCD

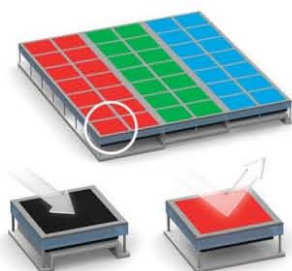
Liquid in each of three layers has a molecular structure that matches a different color of light.

PRO Layers go transparent and can be stacked.

CON Inefficiencies limit overall brightness.

WHO Kent Displays

STATUS Available in Japan



MIRASOL

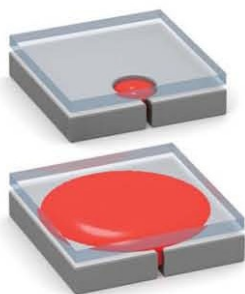
A MEMS device moves a membrane to and from a stack of optical films, changing the wavelength of the light reflected.

PRO Has crisp, fast video, low power consumption, and visibility in sunlight.

CON Is expensive to produce; white is a challenge.

WHO Qualcomm

STATUS Available in small screen sizes



ELECTROFLUIDIC PIXELS

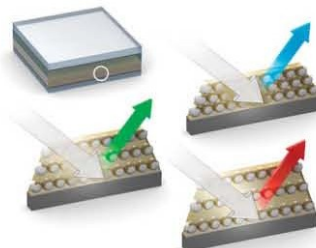
Voltage pulls an inklike fluid out of a small reservoir and into view.

PRO Materials and mechanism are similar to ink on paper; switches fast enough for video.

CON Coming late to the race.

WHO Gamma-Dynamics/University of Cincinnati

STATUS In laboratory



PHOTONIC INK

Electrically active polymers between nanobeads swell or shrink to change the size of the optical cavity and therefore the color.

PRO A single pixel can generate any color.

CON To date, brightness is a problem, and devices have limited lifetimes.

WHO Opalux

STATUS In laboratory

EMILY COOPER

Want color? The current approach is to add a red-green-blue color filter array over the pixels. The problem is that this reduces the brightness by a factor of three, because each primary color filter passes through only one-third of the visible spectrum of light. So, at each color pixel the display can reflect only 10 to 15 percent of the available light. The first color electrophoretic displays, expected to reach consumers late in 2010, will use very weak color filters; this will crank up the brightness at the cost of color saturation.

For bright, full-motion color images on portable screens, LCDs dominate. First developed in the early 1970s and almost continuously improved since then, LCDs are hard to beat for almost any characteristic *except* efficiency. That's why Time Inc. recently presented its futuristic concept version of an electronic *Sports Illustrated* on a standard LCD, and Apple's new iPad sticks with this established technology.

LCDs are energy hogs for several reasons. For one, an LCD works by polarization, which means that at least 50 percent of available light is lost because it doesn't pass through the polarizer. It loses more light to color filters, ultimately wasting about 90 percent of the light from its backlight. So the backlight has to be intense, and it saps power, but that's the only way you can get a bright, crisp, vivid image. The upshot is that LCDs convert electricity to viewable light with pitifully low power efficiency—just 2 to 3 percent.

Worse yet, the readability of both LCDs and the newer organic LED displays, which must also rely on electrically generated light, dramatically deteriorates outdoors. The displays simply cannot compete with direct sunlight, which is about a thousand times as bright as typical indoor lighting. Even a slight sunlight reflection is far brighter than the light coming out of an LCD screen.

The final blow against LCD as the ultimate display technology is that for many people, long-term viewing of an LCD strains the eyes. E-paper displays generally don't cause eyestrain because they automatically reflect—literally—the brightness of your surroundings.

So today's e-paper has readability and low power, and LCDs have brilliant colors and full video motion. Is there a technology that can do it all? A few of the contenders

To Market, to Market



AMAZON KINDLE

DISPLAY

Electrophoretic
pixel from E Ink

RELEASED

2007



G-CORE MINI-CADDY

DISPLAY

Mirasol from Qualcomm

RELEASED

2009



QUE PROREADER

DISPLAY

Electrophoretic
pixel from E Ink
on proprietary
plastic display

RELEASED

2010 (planned)

are bistable liquid crystal, cholesteric liquid crystal, microelectromechanical systems (MEMS), electrowetting, and electrofluidic technology, as well as new generations of electrophoretic technology.

These technologies exploit radically different principles and offer varied features. None of them yet provide the ultimate 2020 display experience of low power, readability, bright color, and full-motion video. But at least a few of them are getting close to providing color e-paper that would be as bright as the monochrome Kindle.

FIRST OUT OF THE LAB later this year will be the multimode display, from Pixel Qi, in San Bruno, Calif. This display takes a brute-force approach, combining reflective and transmissive liquid crystal technologies in an attempt to get the best of both worlds. The display, called 3Qi, operates in three different settings: standard color LCD, black-and-white e-paper, and a limited color e-paper mode. If you're using your laptop in bright sunlight, you would manually switch to the e-paper mode, relying on reflected light; in a dark environment you would switch to the backlit LCD. Combining all these features into a single product causes some loss of maximum brightness and color, but the versatility and low power consumption may make it compelling for consumers, at least for now. [See "Pixel Qi's Everywhere Display," *IEEE Spectrum*, January 2010, spectrum.ieee.org/computing/hardware/winner-pixel-qis-everywhere-display.]

Other developers are trying to push LCD technology into a new realm of performance. Reflective displays based on liquid crystals have been around for decades but have so far failed to impress—think of the drab greenish-gray displays on cheap calculators and digital watches. We should be able to do better, and Kent Displays, in Ohio, spun out of the Liquid Crystal Institute at Kent State University, is doing just that. The company's cholesteric liquid crystal molecules have a helical structure (like a spring or DNA). Shine white light on a layer of cholesteric liquid crystal and, in theory, half of the light will have a circular polarization (left or right rotation) that matches up with the liquid crystal. Also, as it travels through the liquid crystal the light encounters a periodically changing

refractive index, so in total it can reflect a little less than half the light associated with a certain color. Like regular LCDs, the cholesteric liquid crystal can be reoriented with the voltage so that the reflectance can be switched on or off.

To produce a full-color display, Kent makes three separate primary-color films of liquid crystals, each with its own electrodes and voltage control. Then the company laminates the three liquid crystal films into a single paper-thin sheet. That means each color can seem relatively bright, because the other color sheets can turn transparent when necessary.

Each layer of film is not perfectly efficient, optically speaking, so with three laminated layers the reflectance is about 30 percent—still far from what we'd want for our ideal display a decade from now. Interestingly, Kent has also demonstrated panels with solar cells integrated beneath the display, so its products can satisfy our 2020 requirement of making battery charging an infrequent inconvenience.

These cholesteric displays have potential uses beyond the e-reader of the future. Kent recently demonstrated color-changing e-skins that can conform to a 3-D surface. Just as the iPhone made the keypad disappear, technology like Kent's might make the entire cellphone case a reconfigurable display.

THINK OF THE WINGS of a blue morpho butterfly. Beautiful, eh? That kind of intensity would certainly satisfy our 2020 display criteria. Back in 1996, inspired by those butterfly wings, MIT launched a spin-out company called Iridigm Display Corp. (from *iridescent* reflection and *paradigm* shift). Qualcomm MEMS Technologies acquired Iridigm in 2004 and tagged the technology Mirasol.

The display exploits a principle called interferometric modulation. Iridescence, like that on a butterfly, relies on the thickness of a resonant microcavity, the thickness being just a fraction of the wavelength of light to be reflected. The physics are similar to those of the cholesteric liquid crystal but without the strong dependence on polarization. The Mirasol display creates these microcavities using a combination of MEMS mirrors and a stack of thin optical film.



READIUS

DISPLAY

Electrophoretic pixel with plastic electronics

RELEASED

Announced 2008; has yet to ship



SONY READER

DISPLAY

Electrophoretic pixel from E Ink

RELEASED

2006



APPLE iPad

DISPLAY

LCD

RELEASED

2010 (planned)

When ambient light hits the structure, the height of the optical microcavity between the MEMS mirrors and the optical film resonates with just a narrow set of wavelengths of light (a single color), and the mirrors reflect only that color. Pixels have preset mirror heights that reflect red, green, and blue light, with multiple mirrors arranged side by side. Apply voltage and the MEMS mirrors move closer to the optical film, the microcavity disappears, and the pixel reflection turns black.

This scheme can easily produce a pixel with intense color, but making a pixel that switches between more than two colors is a challenge. In displays, manufacturing has to be ultrasimple to keep costs low. Think about the \$200 you pay for a 24-inch LCD monitor; displays, unlike tiny microprocessors, must be really inexpensive per unit of area. Right now, it isn't economically feasible to make MEMS pixels that switch between numerous microcavity heights.

Today these full-color displays reflect about 25 percent of ambient light—much better than full-color electrophoretic displays but not yet approaching our 2020 goal of the 80 percent reflectance of paper. Like electrophoretic displays, they're viewable in sunlight and they're bistable, with the huge power savings advantage that provides; they also can be rapidly refreshed in microseconds, allowing full-motion video. The speed is fast because the mirrors need to move only a very short distance—just hundreds of nanometers.

NATURE HAS OTHER LESSONS for designers of color screens. The adaptive, color-changing skin of chameleons and bobtail squid is biologically complex. But optically, it's easy to understand. The skin contains pigments that concentrate into small dots when the muscle relaxes. Transparent muscle fibers stretch out the skin, thereby enlarging the pigments so that their color becomes visible. When the muscle fibers relax, the pigments spring back into small, barely visible dots.

Conventional printed media exploit a similar principle to mix colors. In theory, a printed image can be made from dots of cyan, yellow, and magenta that overlap to create the full spectrum of color. Using smaller dots, wider spacing between dots, or no dots at all simply *Continued on page 54*

Academic Writing Task 2

Task 1	150 words	20 minutes	Summary of information
Task 2	250 words	40 minutes	Discursive essay

What do I have to do in Task 2?

Task 2 is a topic on which you have to write a discursive essay. The topic may be in the form of a statement or a question. Sometimes different or opposing views are expressed; sometimes there is one view to discuss.



CONTENT

You must answer all parts of the task. You need to make your own position clear and provide main ideas and supporting arguments to illustrate this. You should write a clear introduction and conclusion.

How can I make sure I answer all parts of the task?

You should analyse the task carefully so that you know exactly what you have to write about.

Look at the notes on task A and the summary of these in the table below.

A

- *Some people think that teenagers' use of the internet should be limited. Others feel that the internet is an academic resource that they should have free access to, in order to do things such as homework and projects.*
- *Discuss both these views and give your opinion.*

such as tells me that these are examples – I can discuss other things if I want to.

I must say what I think.

These are opposing views.

I must discuss both views – so it would help to think about who would have these different views and why.

Task	Are opposing views expressed?	What are the key words?	How many parts must I write about?
A	Yes	teenagers / internet / limited / academic resource	Two – limiting or not limiting internet use
B			
C			

1 Read these tasks and then complete the table on page 70.

B

Some parents believe that extra private lessons outside school hours, where students work alone with a teacher, can help them do better at school. Others disagree.

What are the advantages and disadvantages of private tuition?

C

Traffic congestion seems to be increasing.

What do you think are the causes of traffic congestion and what, if anything, can be done to reduce the problems?

What is my 'position'?

Your position is your view on the topic. Make sure that you say what your position is, and that it stays the same, i.e. you don't contradict yourself.

Look at task B above. What is your position?

- i Private tuition has more advantages than disadvantages.
- ii Private tuition has more disadvantages than advantages.
- iii The advantages and disadvantages of private tuition are fairly equal.

2 How is task C different from B? What is your position likely to be?

How do I make my position clear?

You should state your position clearly, perhaps as part of your introduction, support it throughout your answer and re-state it (in a different way) in the conclusion.

3 Underline or highlight the writer's position in this introduction to task B. Is it position i, ii or iii?

In many countries students have to compete to get into colleges and universities when they leave school. For this reason, some parents decide to pay for extra lessons to help their children be more successful. On the whole, I feel that this is a good idea, despite some of the drawbacks of private tuition.

4 Change the last sentence so that the paragraph expresses position ii.

5 You need to re-state your position in the conclusion, by pulling together your main ideas and showing how they support your argument. Underline or highlight the writer's position in this conclusion.

Evidently private tuition is something that has to be considered carefully. However, there is no doubt that it can be enormously helpful in preparing students for important examinations by giving them the extra help they need. Overall, these benefits outweigh the disadvantages.

6 Write your own introduction to task B.

What is a main idea?

A main idea is a key point or argument that relates directly to the question and to your position. You only need a few main ideas, but remember that you may need main ideas on both sides of an argument.

7 Complete these notes, which give some main ideas for task B on page 71.

ADVANTAGES

Lessons go at student's
..... of learning

Student can ask
more

More attention
on

DISADVANTAGES

Cost

Students are already
very

Time needed to travel
to

What if I don't have any ideas?

If you cannot think of your own ideas, think about what you have read on the topic in books or magazines, or seen on television.

How do I make my main ideas clear?

Your main ideas should come between the introduction and conclusion, and form the body of your answer. Each main idea should be in a separate paragraph.

What are supporting arguments?

Supporting arguments add extra information to your main ideas. You should link the main idea to the topic and then support it.

8 Underline or highlight the sentence which contains the supporting argument in this paragraph from task B on page 71.

One of the reasons why private tuition leads to better exam results is the fact that the tutor can teach at the student's own pace. This is not possible in a classroom with a lot of students because there, the teacher has to go at an average pace to suit everyone.

9 Note some main ideas for task C on page 71, using these headings.

Causes of congestion

.....
.....
.....

How to reduce problems

.....
.....
.....

- 10 Look again at task C on page 71. Write an introductory paragraph, and another that includes a main idea and supporting arguments.



ORGANISATION

Your answer needs to develop logically from your introduction through several paragraphs to your conclusion. Within the paragraphs, your ideas should be linked together well.

How do I decide how many paragraphs to write?

Write between five and seven paragraphs: your introductory paragraph, three to five paragraphs for the main body of your answer, and your concluding paragraph. Aim to have one main idea in each paragraph.

- 1 This long paragraph would be better if it was broken into two paragraphs. Where could you start the second paragraph? To help you, underline or highlight the two sentences which contain the main ideas.

Private tuition can result in more successful learning for a number of reasons. The most significant of these is the fact that a personal tutor is able to teach individual students at their own pace. A class teacher, on the other hand, has to keep everyone involved in the lesson. This means choosing a pace that suits the 'average' student but may not suit many individual students. Students can get more personal attention when they are taught on their own. They do not have to worry about understanding something straight away, as it can be repeated as many times as necessary and they can ask lots of questions. This is often not practical in a classroom situation because other students may get bored and, as a result, become disruptive.

How can I develop my answer logically?

You need to start each new paragraph with a word or phrase that shows that you are making a new or related point, e.g. *While this is a popular view; Not everyone takes such an approach; Another possible cause; As far as X is concerned.* You should do this to make your ideas clear.

- 2 Which of the following expressions could you use to begin your second paragraph above, so that it links well to the first paragraph?
- i Nevertheless, some people feel that
 - ii It is also the case that
 - iii Initially
 - iv However
- 3 Why are the other three expressions not appropriate?

- 4 Complete the paragraph openers below for the task on traffic congestion (task C on page 71). Avoid using the words *first*, *second* or *third*.

There are a number of ways that we can help solve the problems of traffic congestion.

Paragraph 1 would be to make sure that every family only has one car.

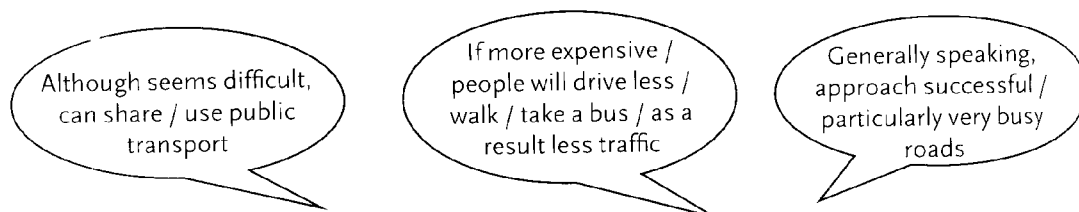
Paragraph 2 would be to increase the cost of petrol.

Paragraph 3 A solution would be to charge people for road use.

How can I link my ideas within paragraphs?

You can link your ideas by using linking words and phrases, e.g. *however*, *yet*, *unfortunately*, *indeed*, *then* or *generally speaking*. Note that they don't always have to be at the start of a sentence.

- 5 Here are some supporting points for the three paragraphs *about traffic congestion*. Add each point to the right paragraph by using the words given to link up the ideas.



- 6 Read task D, then look at the introductory paragraph in which the reference words have been highlighted. Complete the table showing what they refer to.

D

Water is an increasingly valuable resource, but people continue to waste a lot of it. Some governments want to impose permanent water restrictions on domestic and agricultural use. Others feel we should put more effort into recycling water.

To what extent do you agree with these two solutions?

Water is definitely an invaluable resource. Without it, we cannot survive. Today many governments recognise that they need to limit the water that their citizens use. Some also attempt to recycle water. Both approaches to water conservation are necessary and should be promoted, though I feel the first is generally more successful.

it	water	some
they		both approaches
their		the first

- 7 Complete the gaps in the two paragraphs below with a correct reference word from the box. Some of the words will be used more than once.

which
their
these
they
this
the
where

In countries (a) water seems to be readily available, people may, at first, be reluctant to reduce (b) water consumption. So initially, governments need to make (c) citizens aware of the consequences of using too much of (d) valuable resource. Once people realise that water supplies are limited and that (e) have a responsibility for conserving water, (f) task will be easier.

It must be remembered that people use water for many different purposes, (g) *range from running domestic appliances such as washing machines to large-scale agricultural projects that need large quantities of water for irrigation.* In (h) efforts to reduce water use, governments need to target all (i) different types of water consumption. (j) will often involve creating special laws.

- 8 Underline or highlight the other linking words in these two paragraphs.
- 9 Write a third paragraph about recycling water. Include one main idea and some supporting arguments, and link your ideas together well.

VOCABULARY

You need to show that you have a range of vocabulary related to the topic and that you can use these words appropriately and accurately in your answer.

How can I improve my vocabulary range?

You have to know enough words to be accurate and avoid repetition. You can improve your vocabulary related to different topics by reading newspaper and magazine articles and noting some of the topic vocabulary.

- 1 Read the extract below, in which the vocabulary related to task C on page 71 has been highlighted. What sort of publication do you think it comes from?

CLUNK, CLICK, VROOM – AND AWAY WE GO. Every day, millions of us climb into our cars and set off on journeys to work, to the shops or just to enjoy ourselves. And once inside our cars, few of us are inclined to spare a thought for the environmental impact of driving in heavy traffic. Advertising consistently portrays cars as symbols of personal status and freedom, and sources of comfort and convenience.

But behind the shiny commercials, the costs of

our car-dependent lifestyles are becoming increasingly serious. The lengthening traffic jams, demands for new roads, increasing air pollution and threat of climate change are all issues we must tackle sooner rather than later.

Emissions from different forms of transport are the fastest-growing source of greenhouse-gas pollution – mainly in the form of CO₂ arising from the combustion of diesel and petrol.

- 2 Complete these paragraphs for task C, using the highlighted words or phrases in the extract on page 75.

Most people would agree that traffic problems are increasing worldwide. In many large cities, it is hard to drive freely because traffic jams are so common. As a result, (a) is now a serious problem in cities because so much (b) is being used, and the issue of (c) has been directly linked to the human need for fast methods of (d) Why has this happened?

Initially, cars were a practical way of getting from A to B. They were not built to travel at high speeds and

dual carriageways were unknown. If their (e) were relatively short, people often chose to walk, rather than drive. At this time, governments responded to complaints about (f) congestion by building (g) , unaware of the (h) this might have on the environment.

Nowadays, cars have become (i) - everyone wants one and it's hard to stop this because of their (j) and Unfortunately, we have become used to our (k) and we are reluctant to change.

How can I improve my accuracy?

You need to pay attention to how you choose, form and spell words. You will lose marks if you make mistakes in these areas.

- 3 Complete the gaps in these paragraphs with the correct form of the words in the box.

crowd
special
drive
delay

waste

few

increase
manufacture
expense
charge

Beijing is a very crowded city and traffic jams are common, (a) at peak travel times. Between six and seven in the evening, (b) know that the traffic will be bad and that they will have to expect (c) on their journeys. Everyone has got used to this, although no-one likes (d) time stuck in traffic.

In the past, there were far (e) cars in Beijing because they were too expensive to buy, but nowadays an (f) number of citizens can afford one because the car (g) industry in China is booming. In addition to this, petrol is relatively (h) compared to the prices (i) in many other countries.



GRAMMAR

You need to show that you can write a range of sentence types and that you can use grammar accurately. You also need to punctuate your writing well.

How can I show a range of sentence types?

You should include both simple and complex sentences in your essay. (Complex sentences contain more than one clause.)

Look at this paragraph from a student's essay. The sentences are all simple, so the examiner cannot give a high mark for grammar, even though the meaning is clear.

Nearly all countries have traffic problems. They can be hard to solve. Local people can reduce some of the problems. They can choose to walk rather than drive. But this is often not a popular option. So the number of vehicles on the roads rises. However, sometimes there are poor road or traffic conditions. There is not much the public can do about this. Governments must take steps to reduce congestion. This means imposing laws.

Here is the same paragraph, re-written with a wider range of sentence types. This will get a better mark.

Nearly all countries have traffic problems, which can be hard to solve. Local people can reduce some of the problems by choosing to walk rather than drive, but this is often not a popular option. So the number of vehicles on the roads rises. If there are poor roads or traffic conditions, however, there is not much the public can do. Either way, governments clearly need to take steps to reduce congestion and this may mean imposing laws.

How can I improve my accuracy?

As well as checking for grammar mistakes, you should also make sure your punctuation is accurate.

- 1 Find the punctuation errors in this paragraph (there is one on each line).

commas needed round *unlike bicycles*

- a
- b
- c
- d
- e
- f
- g

It is a well-known fact that cars and buses unlike bicycles use lots of petrol and create a great deal of pollution, surely something can be done about this. If we cannot get people to walk or, share vehicles we should put more pressure on scientist's to build solar powered engines. Although it may take some time to achieve this, it would be worth it? There are other alternatives, too. For example: if we all started driving electric cars, the world would be a much cleaner place

ACTION PLAN

- ▶ Analyse the task to see how many parts you have to write about.
- ▶ Decide on your position and your main ideas.
- ▶ Introduce your answer by re-phrasing the question and stating your position.
- ▶ Write three to five paragraphs on your main ideas, with supporting arguments.
- ▶ Link your ideas together so that your answer is logical and clearly developed.
- ▶ Try to use a range of relevant vocabulary and sentence types.
- ▶ Conclude by re-stating your position and summing up your arguments.
- ▶ Check your answer for errors and count the number of words you have used.

8 → ANSWERS PAGE 112
PRACTICE TEST PAGE 105