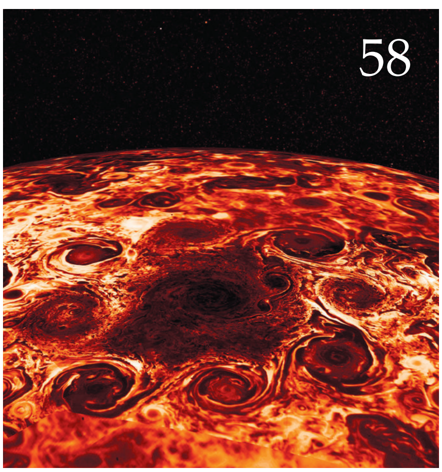
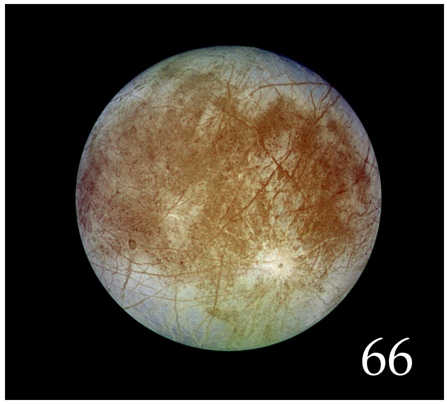
AMERICAN
Scientist

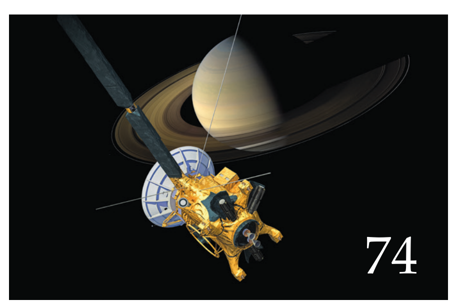
A TOUR OF THE SOLAR SYSTEM

AMERICAN Scientist

Special Collection







Cover Illustration Credit *NASA/JPL*

From the Center to the Edges

Over the years, *American Scientist* has been honored with articles from authors who are science leads on numerous NASA missions. For instance, in 1972, the magazine featured articles by Apollo 15 astronaut Joseph Allen on that Moon mission, and rocket pioneer Wernher von Braun on the then-upcoming Space Shuttle program. In the magazine's more recent history, going back to 2000, we've had articles that feature every planet in the Solar System, plus a number of other bodies orbiting our Sun, authored by prominent astrophysicists, so we have compiled this focused collection that looks at our home system from the Sun at the center to its outermost reaches.

We are delighted that this collection has an article by planetary geologist Paul K. Byrne about Venus. The science in this area moves so quickly that Byrne has already reported additional findings on the *American Scientist* Long View blog: he and his colleagues used image analysis and modeling to show that Venus has plate tectonics, but that the plates look quite different from those on Earth. Space physicist Daniel N. Baker wrote for the magazine about the Van Allen radiation belts that surround Earth—belts that are named for James A. Van Allen, who was Baker's PhD advisor in the 1970s. Physicist Walter Goetz of the Max Planck Institute for Solar System Research was on the team that developed the robotic arm camera for the Phoenix mission to Mars, and he details that mission's results. Scott Bolton, the principal investigator for the Juno mission, gave us the latest updates about Jupiter. Matthew S. Tiscareno on the Cassini mission's imaging team takes us through some amazing images of Saturn. (An earlier article on Saturn by Cassini imaging team lead Carolyn C. Porco is available on our website.) And S. Alan Stern, principal investigator of the New Horizons mission, tells us what that mission found out about Pluto and beyond. Stamatios M. Krimigis and Robert B. Decker, principal co-investigators on the Low Energy Charged Particle Experiment on the Voyager 1 and Voyager 2 spacecraft, detail what those missions found on their long journey through our Solar System and past it. We also have first person interviews with Gene Parker, for whom the Parker Solar Probe was named, and with Dante Lauretta, principal investigator on the OSIRIS-Rex mission that returned samples from an asteroid.

That's not all this issue covers. We take a look at why the two faces of our Moon are so different, how it rains diamonds on Neptune and Uranus, what causes tides on Europa, and new discoveries about mountains on Mercury, to name a few. We also cover the formation of the Solar System itself.

Whether you are receiving this collection as a new subscriber, or you have been a long-standing reader of *American Scientist*, we thank you for supporting nonprofit science communication. Your support allows the magazine to bring you cutting-edge reports directly from experts on the front lines of research, across all fields of science. Feature articles in *American Scientist* are written by scientists and engineers about their own peer-reviewed research, tailored to provide readers with an understandable and engaging account not only of the authors' results, but also of their research journeys.

As a subscriber, you also have access to digital versions of past issues of *American Scientist* on our website. PDFs of past articles are available to subscribers at a reduced rate (and are free to download for Sigma Xi members). *American Scientist* has been in publication since 1913, so there's a long history of past articles to draw on. Many more blogs, podcasts, and other multimedia are available on the *American Scientist* website as well.

We hope that you find this collection useful and informative, and we hope that you will continue to be a subscriber to *American Scientist* for years to come!

Fenella Saunders
Editor-in-Chief

Sculpting Our Planetary System

The discovery of thousands of planets orbiting other stars has given us three surprising insights about our Solar System. First, we are weird: Our Solar System is a 1-in-2,000 rarity. Second, planet formation is a dynamic process, characterized by large-scale orbital drift as well as violent collisions and the ejection of young planets into interstellar space. Lastly, the second point may explain the first one—that is, how our Solar System formed is likely the root cause of our weirdness.

My job is to understand how planets form. This is no easy task, given that we live in such an oddball system. But computer models are improving their ability to realistically simulate the growth of planetary systems. A picture is emerging—albeit a still incomplete one—of how planets form in a larger, galactic context. Simulations show that small divergences at key stages in planets' growth can lead to dramatic differences in fully formed systems. Our system's structure holds clues about which path it followed at critical junctions.

Within disks of gas and dust swirling around young stars, planetary researchers think the story goes as follows: Dust grains grow into pebbles. Pebbles drift through the disk and clump into 100-kilometer-scale *planetesimals,* or asteroidlike objects. Planetesimals and more drifting pebbles then grow into planetary embryos with a mass equal to or greater than that of Mars.

A massive embryo launches density waves that exchange angular momentum with the gaseous disk, causing the embryo's orbit to shrink (or, more rarely, to grow) in a process called *migration*. Some particularly large embryos capture great quantities of gas and become giant planets. A few million years later, the gas in the disk dissipates, which in most systems triggers a phase of instability. Growing rocky planets may undergo giant collisions, such as the impact that formed our Moon. Instabilities among giant planets are even more extreme, usually culminating in the ejection of one or more planets into interstellar space.

To fit all the pieces together, we need to understand the interconnections between the different phases of planetary growth, at various times and places and on different size scales. For a long time, astronomers assumed that planets just plopped together in place; now we know that is not even close to being true. We think that the key processes just described—orbital migration and dynamical instability—are the disruptive architects of planetary systems.

Taxonomy of Planetary Systems

When we survey the population of exoplanets, super-Earths (with masses that fall between that of Earth and that of Neptune) are the most common type we see: About half of all stars have a super-Earth orbiting them, circling them more closely than Mercury does the Sun. The equivalent region of our Solar System is completely empty. How did all those super-Earths get there?

Migration generally causes the orbits of planetary embryos to shrink, so it is natural to imagine that migration may have played a role. Let's assume that the building blocks of super-Earth systems are similar to those that formed the Solar System, with small, rocky embryos populating the inner region and large embryos forming past the *snow line*, the boundary beyond which temperatures are cold enough that ice could be used as a building block. Large embryos might migrate inward to the inner edge of the disk, which acts as a migration trap.

Our models show that successive migrating embryos do not collide but are driven into orbital *resonances*, in which the orbital periods of adjacent planets form ratios of small integers. For example, in 3:2 resonance the outer planet completes two orbits in the time it takes the inner planet to complete three. Migration naturally generates chains in which each pair of neighboring planets is in resonance, stabilized in part by the gaseous disk. When the gas dissipates, the chains break, but sometimes the or-

bital patterns persist after the gas disappears. Several examples of such resonant systems have been discovered, such as the TRAPPIST-1 system, which hosts seven tightly spaced Earth-sized planets.

The "breaking the (resonant) chains" model matches the measured properties of the super-Earth population. Giant exoplanets appear to fit this model as well. They are commonly found on stretched-out, eccentric orbits. In these systems, multiple gas giants probably originally formed in circular orbits. Embedded in their gaseous disks, they migrated into resonances. Upon dispersal of the gas disk, the resonant systems became unstable, leading to planet-planet scattering and the ejection of one or more planets. The surviving planets ended up on eccentric orbits, the scars from their systems' violent pasts.

Scars of the Solar System

Although our Solar System is unusual, it bears its own scars from a violent past. The configuration of the familiar planets makes more sense when seen in the context of these new models.

Several lines of evidence point to an instability in the orbits of the giant planets, which would have been naturally triggered if they had migrated into a resonant chain. For decades, studies of planet formation were plagued by the *small Mars problem*. The classical model of planet formation assumes that the terrestrial planets (Mercury, Venus, Earth, and Mars) grew from a disk of planetary embryos and planetesimals that extended from Mercury's orbit out to Jupiter's. Yet simulations of that model systematically produce a Mars that is as massive as Earth, an order of magnitude larger than the real Mars.

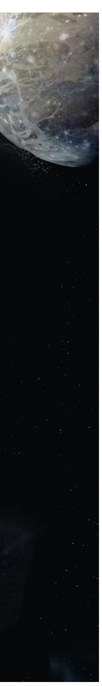
Three solutions have been proposed to the small Mars problem. The *Grand Tack* model, developed in 2011 by my colleagues and I at the Observatoire de la Côte d'Azur in Nice and the Laboratoire d'Astrophysique de Bordeaux, proposes that Jupiter migrated inward and then outward, clearing much of the rocky material from the Mars region. The asteroid belt was emptied and then refilled by the migrating Jupiter. The *Low-mass Asteroid Belt* model, spearheaded by researchers at the University of California, Los Angeles, and the University of Zurich, assumes that planetesimals formed in a narrow ring between the present-day orbits of Venus and Earth, with almost none in the asteroid belt. In this model Mars represents an embryo kicked out of the ring and starved. Finally, the *Early Instability* model, developed in 2018 at the University of Oklahoma, holds that the giant planets' instability happened shortly after dispersal of the gaseous disk, depleting the Mars region but not the Earth-Venus zone.

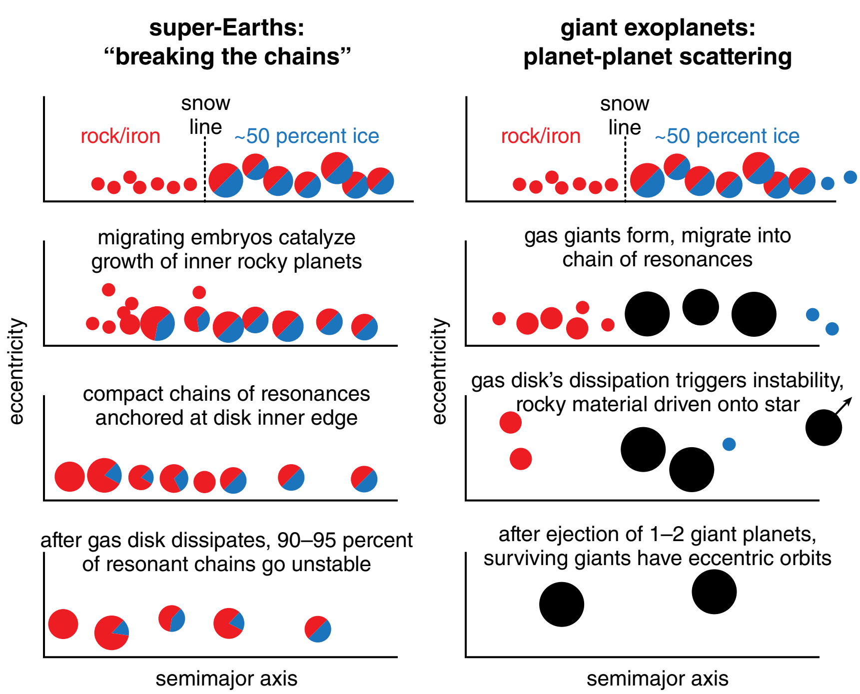
Each model matches the terrestrial planets' masses, orbits, and inferred formation timescales, as well as the observed structure of the asteroid belt. Yet each has a potential Achilles' heel. For the Grand Tack model, it's unclear whether Jupiter's outward migration is viable within realistic disks and formation scenarios. For the Low-mass Asteroid Belt model, it's unclear whether rings of planetesimals really form. For the Early Instability model, we don't know enough about the timing of the Solar System's instability. Using all resources available, from exoplanet studies to meteorite analyses to computer simulations, the next steps are to figure out which of these models may represent our system's true past.

Even now, we can start to piece together the formative events that made our Solar System weird. Jupiter must have grown quickly and starved the inner Solar System, blocking inward-drifting pebbles to prevent the terrestrial embryos from growing massive enough to migrate. Later, Jupiter blocked the inward migration of Uranus and Neptune, which can be thought of as failed super-Earths. But Jupiter's migration cannot have brought it too close to the terrestrial planet region, or else Earth's growth would have been violently disrupted; perhaps, as in the Grand Tack model, Jupiter was held back by Saturn's presence. Similarly, Jupiter's instability could not have been too strong. According to our models, the instability involved strong gravitational scattering between Jupiter and an ice giant (a sibling of Uranus and Neptune that later was ejected), but Saturn and Jupiter never scattered off of each other. Earth would not have survived if they had.

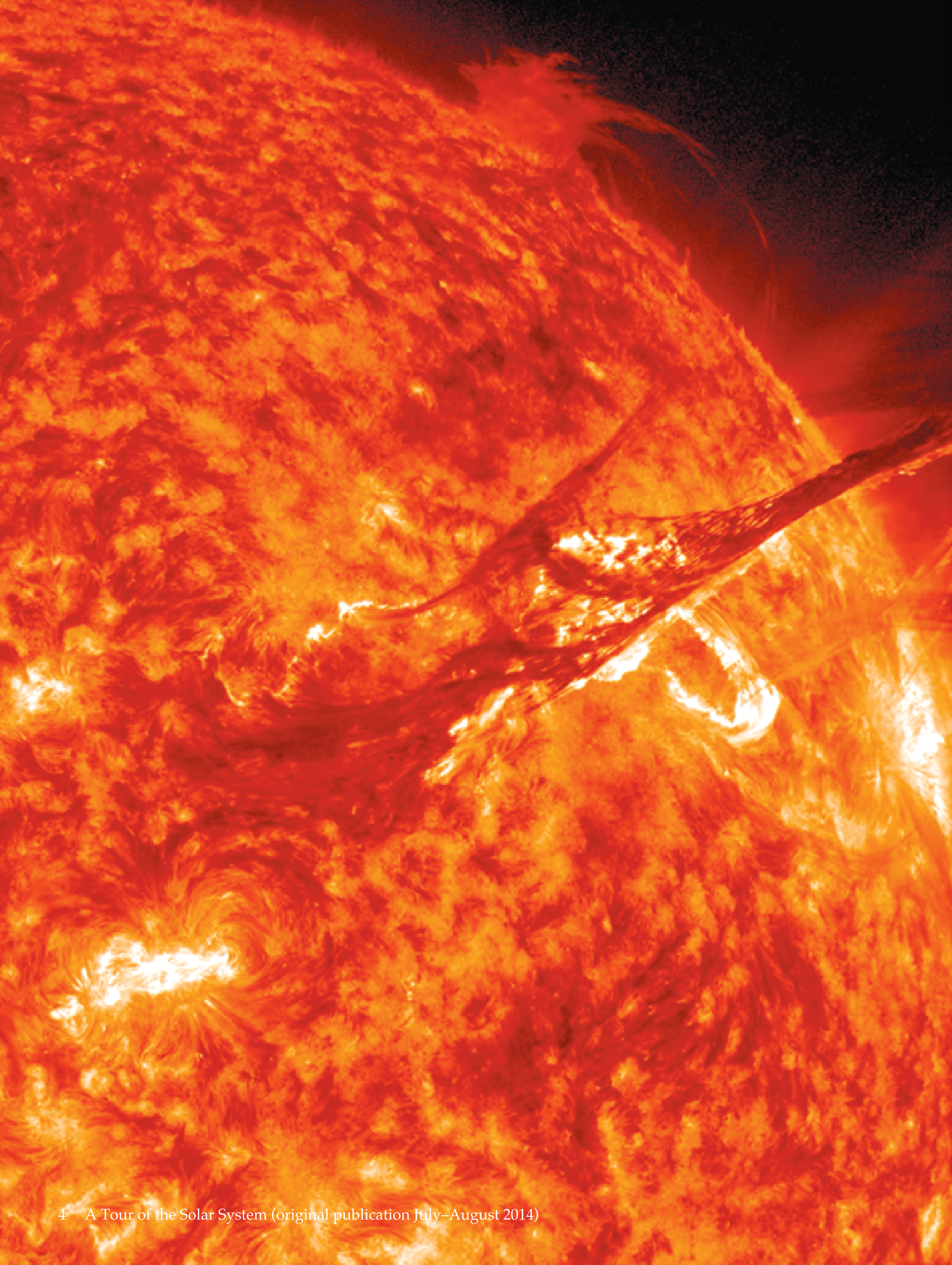
What all the models tell us is that planetary formation is a dynamic process, marked by big instabilities and lurching movements. No wonder there is such a dizzying array of systems out there. We've long wondered whether the Earth is normal or an oddball. The answer is, it is both. The processes shaping our Solar System were commonplace, but the outcome was unusual. Still unknown: Is the Solar System unique?

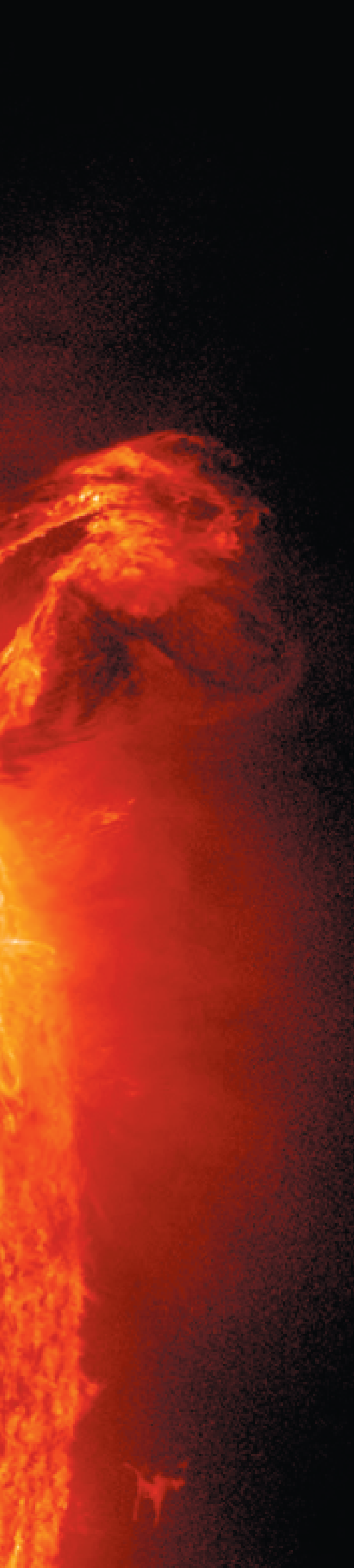
Planets may undergo giant collisions, such as the impact that formed our Moon (*left*). Theories of how orbital migration and dynamical instability shape populations of exoplanets include the "breaking the chains" model (*below left*) and the "planet-planet scattering" model (*below right*). (Illustration at left courtesy of Hagai Perets, Technion; below courtesy of the author.)





Sean Raymond is an astrophysicist studying planet formation at the Laboratoire d'Astrophysique de Bordeaux, France. Email: rayray.sean@gmail.com





The Real Sun, Unmasked

The Solar Dynamics Observatory produces stunning images while investigating the origins of space storms.

Catherine Clabby

From 150 million kilometers away, the Sun seems like a model of consistency. Day after day the colossal nuclear reactor bathes Earth in steady streams of electromagnetic energy. That radiation keeps the planet's climate in balance and anchors our food webs. But take a closer look and it becomes clear that our star has a hidden side that is more complicated, impulsive, and sometimes downright dangerous.

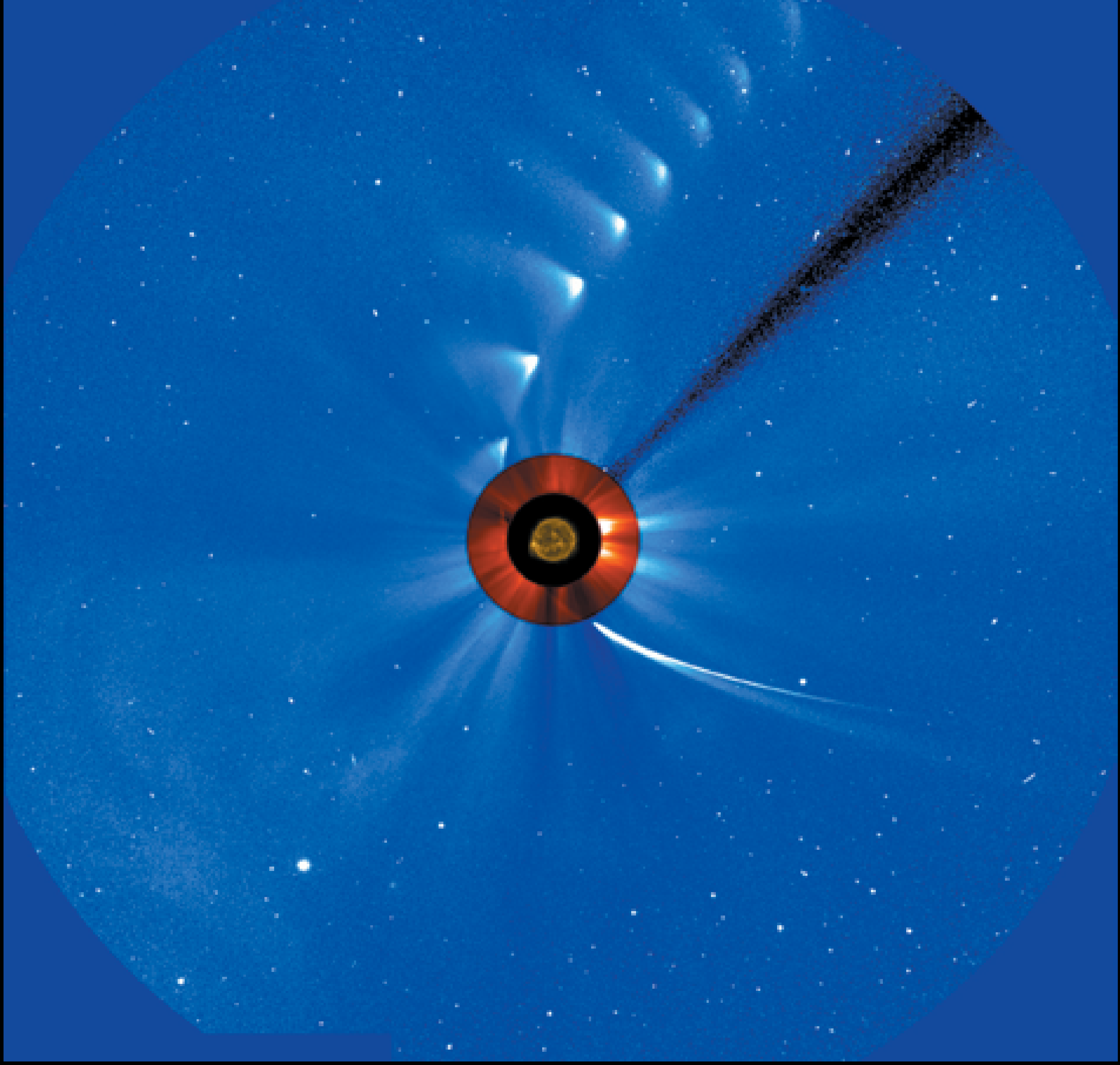
Like all stars, the Sun is highly dynamic. Sometimes it produces powerful magnetic eruptions, including solar flares and coronal mass ejections. These occur on a wide range of scales. In 1859, the most powerful recorded solar storm to hit Earth sent electrical surges sparking down telegraph networks and lit up brilliant aurora displays as far south as Cuba. Today, a geomagnetic disturbance on that scale could cause extensive damage to satellites, GPS systems, power grids, and radio communications. Although estimates are uncertain, the cost could exceed $1 trillion according to a National Research Council report.

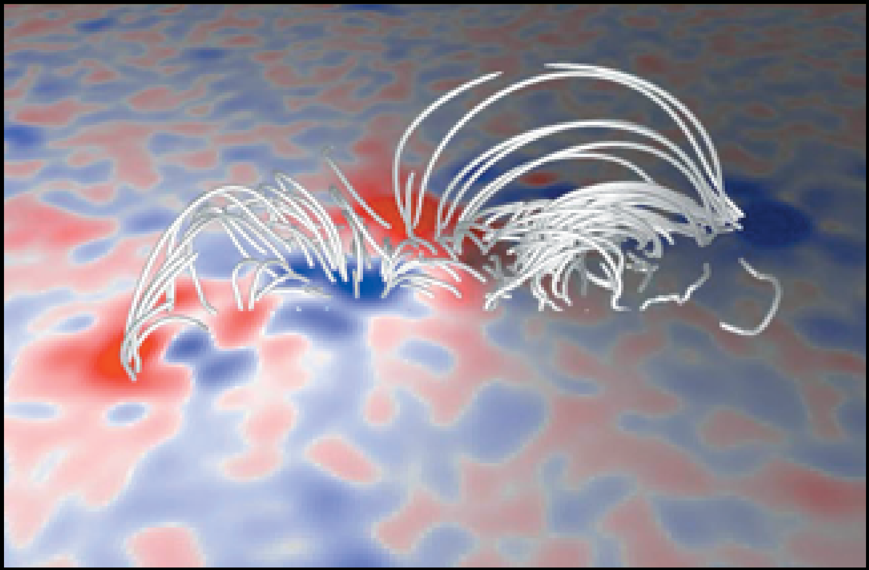
A network of observing stations, in space and on the ground, now watches the Sun across a wide range of electromagnetic wavelengths to monitor its effect on space weather and to provide advance warning when major storms are headed our way. NASA's Solar Dynamics Observatory (SDO) is the newest member of this network. It circles 36,000 kilometers above the ground, in a geosynchronous orbit that keeps it roughly above northern Mexico. From that vantage point, SDO's instruments are producing some of the most stunning images ever made of the Sun and capturing events in real time.

"In the past images were taken every 3 or 15 minutes or half an hour. Now we're taking one every 12 seconds. If you see something happen, you can now go back in the database and see what happened beforehand. That allows you to study the little things that likely influence the big things," says Dean Pesnell, project scientist for SDO, which launched in 2010. The observatory also records the Sun in 13 different wavelengths. Each wavelength reveals different details in the Sun's surface, the photosphere, the overlying chromosphere, and into the corona, the Sun's atmosphere.

Magnetism is key to understanding solar activity, because the Sun is composed not of atoms but of plasma, a brew of positively charged ions and negatively charged electrons. The motion of that electrically conductive plasma—stirred by convection, rotation, and the force of escaping solar winds—creates twisted and variable magnetic fields. When these fields interact, they can unleash the energy that powers flares and mass ejections. "The Sun's magnetic field makes all this

NASA's Space Dynamics Observatory (SDO) captured a long filament of solar plasma erupting from the Sun's corona in August 2012, here in extreme ultraviolet light. With resolution never achieved before, the SDO observes the Sun in 13 spectra every 12 seconds, recording the solar changes that can create space weather.





The speed and direction of comets as they approach the Sun give scientists insights into the shape and strength of the Sun's magnetic fields. In 2013, the European Space Agency and NASA's Solar and Heliospheric Observatory captured the motions of comet ISON in the time-lapse image above. At left, the SDO's helioseismic and magnetic imager creates maps of magnetic fields on the Sun's surface.

happen. We want to know where it comes from. We want to know how it gets to the surface. We want to know how it gets destroyed," says Pesnell. The Sun goes through an 11-year cycle of activity, currently near its peak, but eruptions can happen any time. (See "Reconnecting Magnetic Fields," September–October 2009.)

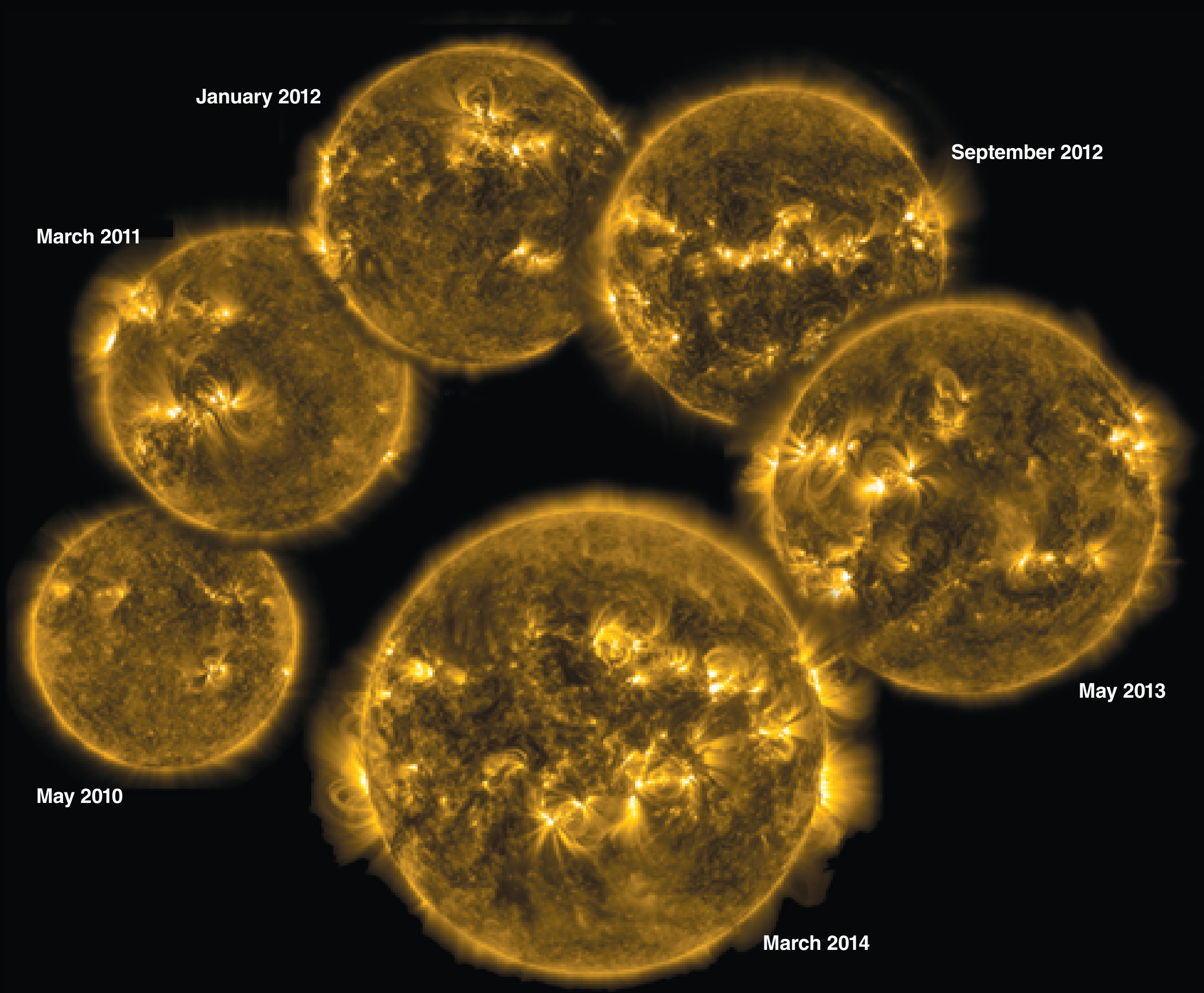
One instrument aboard SDO measures the strength and direction of the Sun's magnetic fields by observing and interpreting how light travels through those fields. Another way to study fields is to study "Sun-grazing" comets that interact with magnetic fields while passing through the Sun's corona. "Magnetic fields in the higher part of the solar atmosphere are very difficult to measure," says C. Alex Young, associate director for science for NASA's heliophysics science division. "By observing this with SDO, we have a new, indirect way to determine the magnetic fields, one of the keys to understanding the causes of space weather." SDO is also monitoring coronal holes, temporary openings in the corona where the solar wind escapes at very high speeds, around 750 kilometers per second. A British team has just reported that fast solar winds seem to increase the number of lightning storms here on Earth. (See "The Origin of the Solar Wind, November–December 2002.)

Although the current solar cycle is quieter than the few that came before, the Sun is hardly calm. In January, a huge solar flare delayed the launch of a private rocket transporting cargo to the International Space Station. Orbital Sciences, the maker of the spacecraft, did not want to risk radiation from the flare damaging the rocket's gyroscopes or avionics. Data from SDO helped the company make that call.

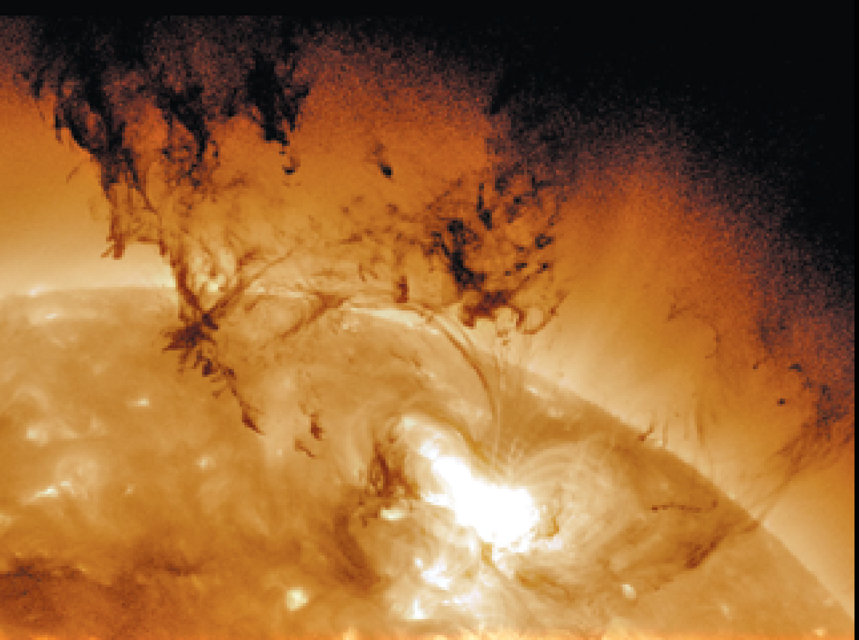
Even better views of how the Sun releases energy are on the way. The Magnetospheric Multiscale mission, to be launched by NASA in 2015, will study how the Earth's magnetic field lines break apart and reconnect, improving our understanding of space weather and solar magnetism in general. A far more daring mission is Solar Probe Plus, which NASA intends to launch by 2018. It will swing within 7 million kilometers of the Sun's surface, far closer than any other spacecraft has managed.

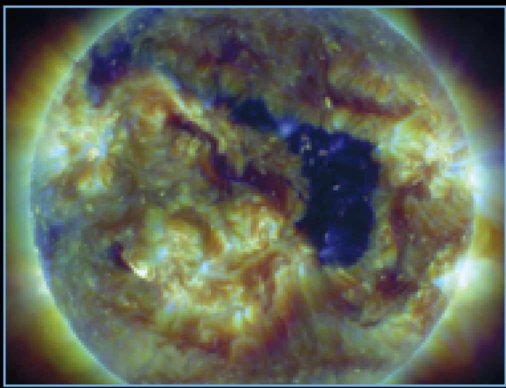
Solar Probe Plus will sample the Sun's atmosphere to better understand how solar particles are energized. It will be our world's first space mission to a star.

Viewing the Sun in different wavelengths allows scientists to see a wide variety of solar materials in different locations, from the surface of the Sun up through its atmosphere. These images were made at the same moment when the Moon moved between SDO telescopes and the Sun. From top to bottom they were observed in 193-angstrom light, 131-angstrom light, and 304-angstrom light.



These images capture increased activity on the surface of the Sun, measured by the number of sunspots visible on the solar disk, between 2010 and 2014. Solar cycles occur roughly every 11 years, reaching a peak in magnetic activity during what's called the solar maximum. The current cycle has displayed the weakest activity ever recorded.



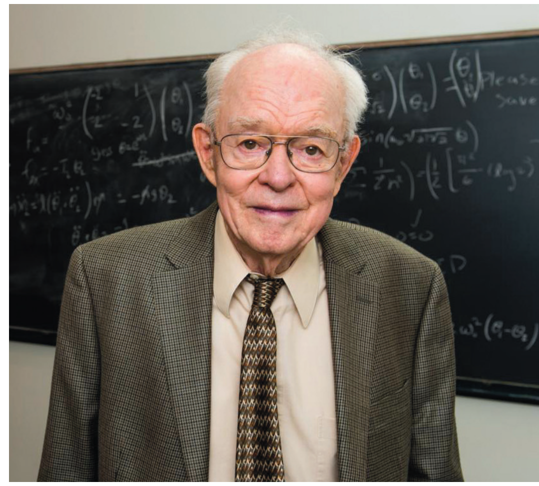


In 2011, SDO recorded a solar flare, the white flash at left, on the surface of the Sun, along with a coronal mass ejection, the darker material, in extreme ultraviolet light. Above, a sizable coronal hole, observed here in three wavelengths of ultraviolet light, rotated towards Earth over several days in 2013. Coronal holes release strong solar wind gusts that carry solar particles to Earth's protective magnetosphere and beyond, sometimes generating aurora.

First Person | Seeking to solve a 60-year-old solar mystery

Parker, Meet Parker

In 1958, astrophysicist Gene Parker mathematically described the solar wind, a continuous flow of charged particles from our Sun that accelerates to supersonic speeds and permeates our Solar System. But 4 years passed before the scientific community accepted that there was anything other than dust between planets. Then in 1962, the world's first interplanetary mission (NASA's Mariner 2) gathered the necessary data to prove Parker right. This year, NASA launched the Parker Solar Probe—NASA's first mission named after a living person—to explore what accelerates the solar wind. American Scientist's *digital managing editor Robert Frederick spoke with Parker, who is now 91, about what he hopes the Parker Solar Probe will discover.*



Jean Lachat/University of Chicago Creative

What do you think kept prior probes from attempting to do what the Parker Solar Probe is going to do, in getting so close to the Sun? Was it technology, or expense, or something else?

Both. It started out when people thought it would be nice to get into the atmosphere of the Sun and see what the conditions are there, what oscillations, what waves, and what disturbances are there. In a very general way, you realize that you could probably discover a lot of new effects if you could just get near the Sun without melting down your spacecraft. Somebody at the Jet Propulsion Laboratory took this challenge seriously and they designed a spacecraft that would survive a short swing past the Sun, reaching a closest approach of four solar radii. That's pretty fantastic. Anything exposed at that temperature is white hot, but with a clever design—a conical shape with the apex pointed toward the Sun—they designed a spacecraft that should stand a swoop past the Sun coming by at four solar radii. But it never got off the ground, because, first, it was expensive. Nobody denied that. Second, and pretty serious, was the objection: What can you learn in four hours? You might see something interesting, but you can't go back and take another look at it. It was realized finally that the best science would be done by compromising. Let's go to 10 solar radii as the spacecraft sweeps around the Sun, rather than shooting for the extreme case of four solar radii. You're sacrificing everything in that extreme case for the survival of the spacecraft. At 10 solar radii, the science opens up. It takes days to go by instead of just hours. You don't have to sacrifice everything. In going by at four solar radii, you cannot expose any instrument to sunlight. You find you're trapped. That's what the present solar probe is all about. The heat is overcome by the heat shield, a big piece of very specially constructed carbon. The analogy the engineers give you is, imagine something made of charcoal, except it's a much more sophisticated and effective form. That's where we stand at the present time. Now we're waiting to see what nature has to show us.

Regarding the broader field of astrophysics, over your long career, have you yourself experienced any surprises? Have you been surprised in the way that your solar wind paper caused others to be surprised?

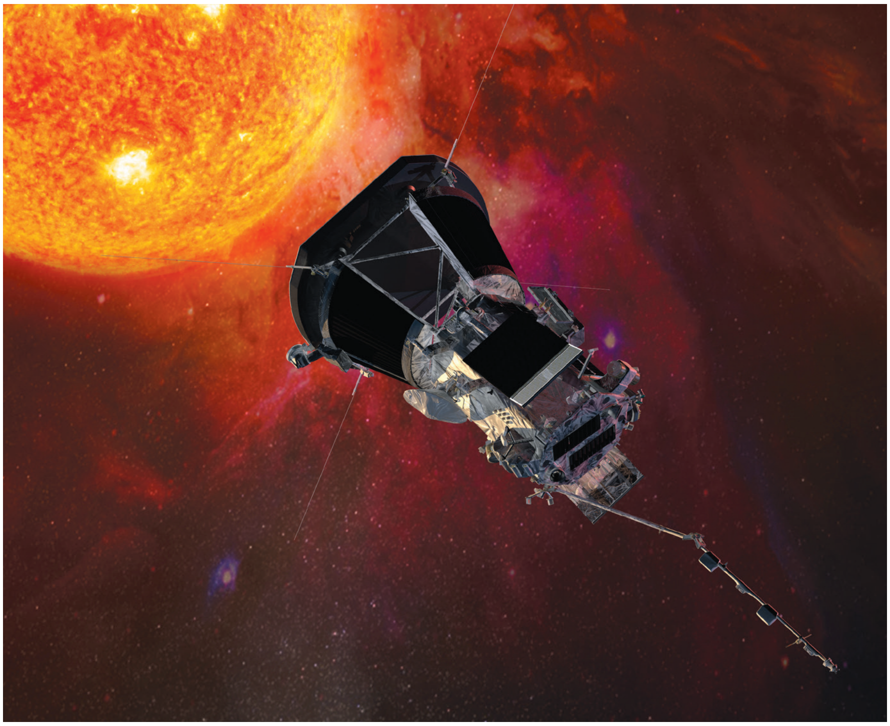
I know a couple of anecdotal examples that are worth a good laugh. I once wrote some papers on the dynamic effects of cosmic rays. You should think of cosmic rays not as particles, but as a continuum again. The dynamic equations are easily written down and solved for various circumstances. It turns out that the galactic magnetic field is inflated by cosmic rays. It's unstable because of the pressure of cosmic rays. And so it's very interesting and quite simple. I wrote it up. It's a cute phenomenon. I sent it in to the *Astrophysical Journal,* and Subrahmanyan Chandrasekhar, who was the editor at the time, sent me a copy of the referee's report. [Chandrasekhar later won the Nobel Prize for Physics and is namesake of the Chandra X-ray Observatory.] The thing that tickled me most is the report. It said, "I had always thought Parker was competent, but . . ." And there followed a long harangue about a single substantive objection to it. When I showed that to the editor he said, "Okay, I'll publish it, if a hostile referee can't come up with a more forceful objection than that." Hostility of that sort is a very common phenomenon, and if you're just a young guy starting out, unknown, it can sometimes ruin your career. People don't know anything about you except this paper, which might be excellent, but it's referred to by an eminent expert as being wrong, so you're wrong and you're dead. I had that happen to me and it didn't really matter, because the paper was right, and rather trivial.

After the solar wind paper, it sounds as though you started in on other problems.

Oh, yeah. The universe is full of problems. We're all mortal, so you'd better get to work on it. You're not gonna last forever.

Have you been involved in any other planned missions to study the Sun?

Only peripherally. I've offered general comments on what it would be nice to do, what we'd like to measure. But the design and building of a spacecraft to check the solar wind is entirely up to technical staff, who never seem to get any credit.



Artist's conception, Johns Hopkins University Applied Physics Laboratory

In order to get seven times closer to the Sun than any previous spacecraft has done, the Parker Solar Probe will use seven Venus flybys over seven years to shrink its closest approach to within 3.83 million miles of the Sun, which is 10 times closer to the Sun than Mercury is.

How did you learn that NASA would name the Parker Solar Probe after you? I'm retired, but I was sitting and working on something or other of no significance. The phone rang, and it was Tom Zurbuchen [Associate Administrator for the Science Mission Directorate at NASA], whom I knew slightly. He said that NASA was proposing to put my name on this spacecraft, but they wanted to make sure it was okay with me. So I didn't have to think too long and hard. I said, yeah, that's a fine idea! He said, thank you, and hung up. The rest is history.

In your heart of hearts, do you most hope that the Parker Solar Probe helps solve those mysteries about the Sun, things you've wondered about for decades? Or do you hope it turns everything on its head, and raises new questions we can't explain at all? Well, there's always the possibility. One can never deny that you may stumble across something that is completely contrary to what we knew. But what's known is already based on general dynamical principles. So I suspect there may well be some surprises, but there isn't going to be some previously unthought-of possibility. I think those are pretty well covered by the available theoretical possibilities. And if the spacecraft wants to prove me wrong, I'm delighted.

An extended audio version of this interview is available online.

Mercury's Mountains

The mysterious rocky planet has the most peaks where its crust is thickest, and more may still be forming.

The Solar System's innermost planet may be hiding big surprises beneath its small, battered surface. One of Mercury's most distinctive features is its long, linear mountain chains, called *lobate scarps*. For years, most scientists interpreted the scarps as wrinkles on a cooling, shrinking, slowly dying world. But a new analysis suggests that lobate scarps may actually be a sign of a hot, churning interior and a surface that remains geologically active to this day.

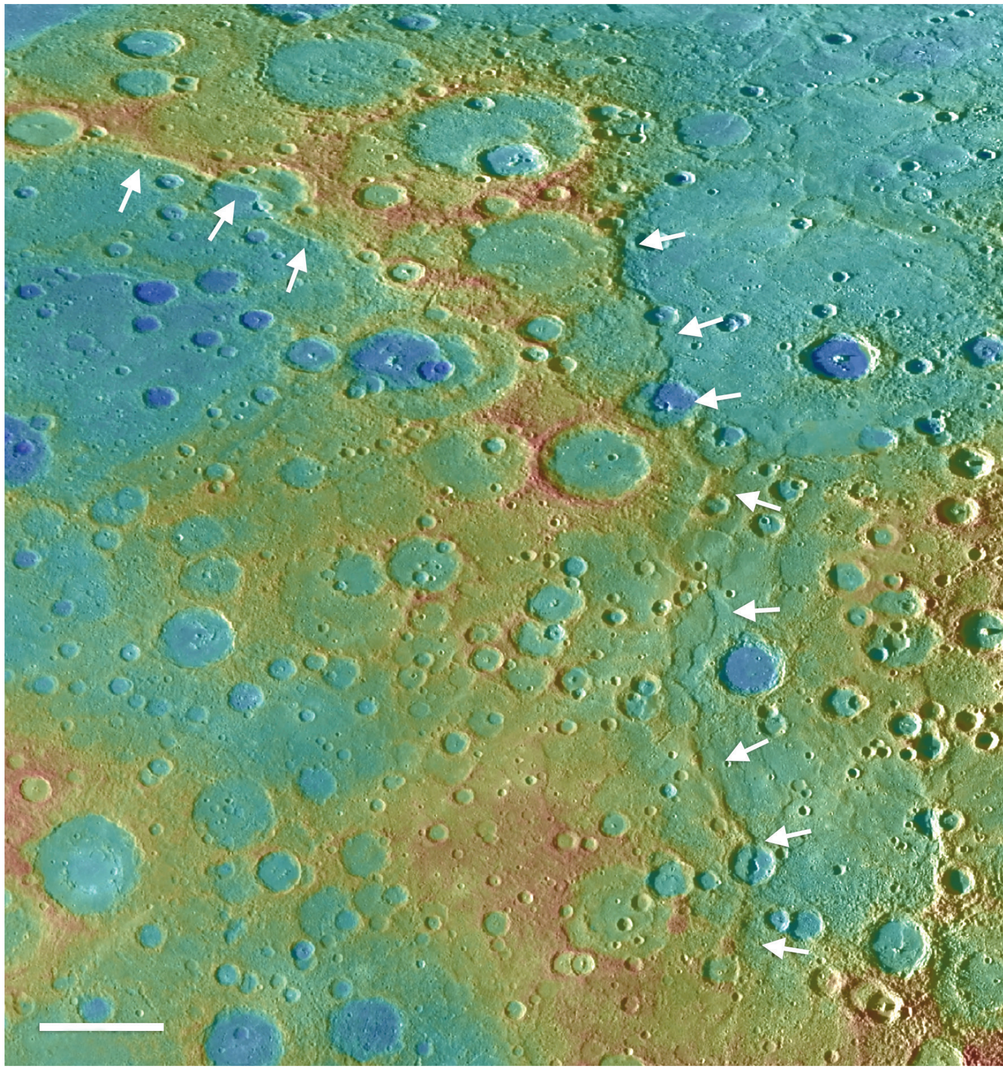
Earth's surface is broken into moving sections, or plates, that create mountains when they collide. However, Mercury's crust is a single continuous shell. Because there are no plates that could crash into one another and form mountains, it was long thought that Mercury's widespread lobate scarps had formed as a result of the planet's interior cooling over time, causing the crust to shrink and randomly wrinkle.

In a paper published in August 2021 in *Geophysical Research Letters*, Thomas Watters and Michelle Selvans of the Smithsonian's National Air and Space Museum and Peter James of Baylor University discovered a pattern in the seemingly random distribution of these mountain chains: Lobate scarps are concentrated in areas of thicker crust, particularly in the southern hemisphere. They also found that these areas of thick crust had been pushed together more forcefully than areas of thinner crust. Although the global distribution of lobate scarps supports the idea that Mercury's surface is indeed wrinkling, their concentration in regions underlain by strained, thick crust suggests that additional factors have influenced their formation. "Something is helping to organize the forces that are acting to produce these faults," said Watters.

The team developed models of Mercury's crustal thickness based on topographic and gravitational data collected by NASA's Mercury Surface, Space Environment, Geochemistry, and Ranging (MESSENGER) mission, which was launched in 2004 and ended in 2015. The models suggest that Mercury's crust was pushed together in specific locations because of geological activity in the mantle, which sits between a planet's crust and core. "There's this phenomenon called downward mantle flow where the material in the mantle is descending toward the core," said Watters. "As it does, it pulls crustal material together, thickens it, and compresses it." This thickening and compression causes the crust to crack and shift, producing lobate scarps.

Very-high-resolution images revealed small, relatively young lobate scarps, suggesting that newly formed faults are still actively modifying Mercury's ancient surface.

Watters believes that Mercury may still be geologically active today because it has a magnetic field, which NASA's Mariner 10 mission discovered in the 1970s. "There's still a hot outer core on Mercury that's liquid, and possibly still moving and convecting to generate that magnetic field," he said. "There's no reason to believe that Mercury has stopped contracting." This idea is supported by photographs that MESSENGER captured during the last 18 months of its mission. Very-high-resolution images revealed small, relatively young lobate scarps, suggesting that newly formed faults are actively modifying Mercury's ancient surface.



Courtesy of NASA/Johns Hopkins University/Carnegie Institution of Washington/Smithsonian Institution

Mercury's mountains are globally distributed but form in clusters (*indicated by white arrows*), suggesting that their formation is not random. Thomas Watters of the National Air and Space Museum and his colleagues used images and mathematical models to understand how crustal thickness may be tied to mountain formation in the absence of plate tectonics. In this false-color photograph, lower elevations are depicted in shades of blue and higher elevations in shades of red.

Studying Mercury up close is not an easy task. Spacecraft require a lot of fuel to stay in orbit because the Sun's gravitational pull is incredibly strong. Once there, spacecraft must then withstand the Sun's scorching heat and intense radiation. Although there have been 48 missions to Mars, including failed attempts, only two missions had been sent to Mercury as of 2017: MESSENGER in the 2000s and Mariner 10 in the 1970s.

In spite of these hurdles, the next voyage to Mercury is already underway. Planetary scientists are eagerly waiting for the BepiColombo spacecraft, jointly developed by the European Space Agency and the Japan Aerospace Exploration Agency, to enter orbit around Mercury in 2025 following its 2018 launch. The spacecraft is named after Italian engineer Giuseppe "Bepi" Colombo, who discovered that Mariner 10 could use Venus's gravity as a slingshot to fly by Mercury multiple times, ultimately allowing it to photograph nearly half of the planet's surface. The BepiColombo spacecraft used this same maneuver to make its first Mercury flyby in October 2021.

Mariner 10 and MESSENGER revealed surprising insights about Mercury's turbulent interior, and BepiColombo is expected to do the same through detailed analyses of the planet's core, chemical composition, and surface. Ultimately, learning more about this enigmatic planet will help inform efforts to understand distant planets outside of our Solar System. "We still have a lot to learn in our own Solar System about how these rocky bodies evolve as they're losing their interior heat," said Watters. "That's going to give us important insight into what we may be finding when we can examine the variety of Earth-like exoplanets that are out there."—*Amanda Rossillo*

Unveiling Earth's Wayward Twin

Venus, the closest planet, seems like a hellish version of our own; studying how it got that way will tell us a lot about the prospects for life among the stars.

Paul K. Byrne

When I started studying planetary geology, I hated Venus. There's no way to see its surface directly, because the planet is wrapped in an unbroken layer of sulfuric acid clouds. Radar can penetrate the murk, but the resulting images are so limited and ambiguous that it's almost impossible to tease apart what exactly is going on. And forget about searching for signs of habitability in that jumble. Even though Venus is nearly the same size as Earth and orbits just slightly closer to the Sun, it is nothing like our planet. Its atmosphere consists of unbreathable carbon dioxide, so dense that it practically flows like an ocean. Temperatures on the ground resemble those inside a self-cleaning oven.

But over time I have come to love the Hell Planet. I now see the complexities that once frustrated me in a different light—as clues to a fascinating, increasingly urgent set of questions.

Some of those questions hit close to home. Venus formed at the same time as Earth, and is presumably made of largely the same materials. How could a world so fundamentally similar to our own have turned out so disastrously different? For a long time, most planetary scientists assumed that Venus went off track in its early days, perhaps right after it formed. Recent studies hint at a different possibility: Venus might have been moderate, even Earth-like, for most of its life before a runaway greenhouse effect transformed it into the infernal pressure cooker it is today. Figuring out which story is correct has major implications for reconstructing Earth's history and for predicting whether a similar process might someday devastate our planet.

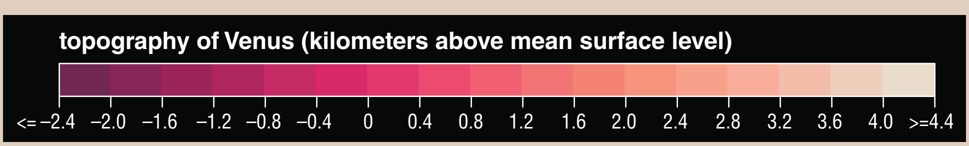
Studying Venus will also tell us a lot about the prospects for finding life around other stars. Exoplanet searches have identified thousands of worlds around other stars, including more than a dozen Earth-sized bodies that could potentially have comfortable temperatures and liquid water. What we cannot tell, yet, is whether any of those places could actually support life. We have neither the theoretical insight nor the observational data to tell whether we live in a galaxy full of balmy Earths, or if we are surrounded by brutal Venuses. We need to get a handle on the fundamental differences between our planet and the one next door if we're to correctly interpret what we see in other planetary systems and correctly target our search for alien life.

QUICK TAKE

Venus is a planetary puzzle. It started out similar to Earth in size, composition, and distance from the Sun, yet ended up with a crushing atmosphere and a dry, lethally hot surface.

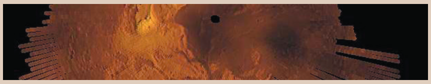
No spacecraft has explored the geology of Venus in more than three decades. A modern mission could investigate why Earth and Venus developed in such starkly different ways.

Earth-sized planets appear to be common around other stars. Venus offers clues about how many of them could actually support life, and about how a habitable planet could die.



Data: NASA & USGS/ Map: © Eleanor Lutz 2017

Venus is as enticing and enigmatic to scientists today as Earth was to the early mariners of centuries or millennia ago. We have named a vast array of features that identify Venus as a world we can explore: Ovda Regio, Niobe Planitia, Metra Corona. But despite its proximity, Venus remains one of the most mysterious planets in the Solar System.



Magellan showed us a world with a vast array of volcanic and tectonic features, including some bizarre structures not seen anywhere else in the Solar System.

Pull quote images courtesy of Donald P. Mitchell/Soviet space program

Inspired by these ideas, I've become a self-professed Venus evangelist. It's not an easy job. Despite its proximity and scientific importance, Venus remains one of the least-explored planets in the Solar System. The last time NASA launched a mission there was three decades ago—three decades during which ever-more-capable spacecraft have visited Pluto, dropped probes into Jupiter, dived through Saturn's rings, and traversed the deserts of Mars. Still, the limited information we do have about Venus provides a strong incentive to learn more. There is evidence of erupting volcanoes, shifting faults, and other geologic activity that shows what Earth might have been like without its oceans. Just in the past few months, radio-telescope studies have revealed tentative hints of peculiar chemistry, and possibly even biological activity, within Venus's clouds (*see box on page 36*).

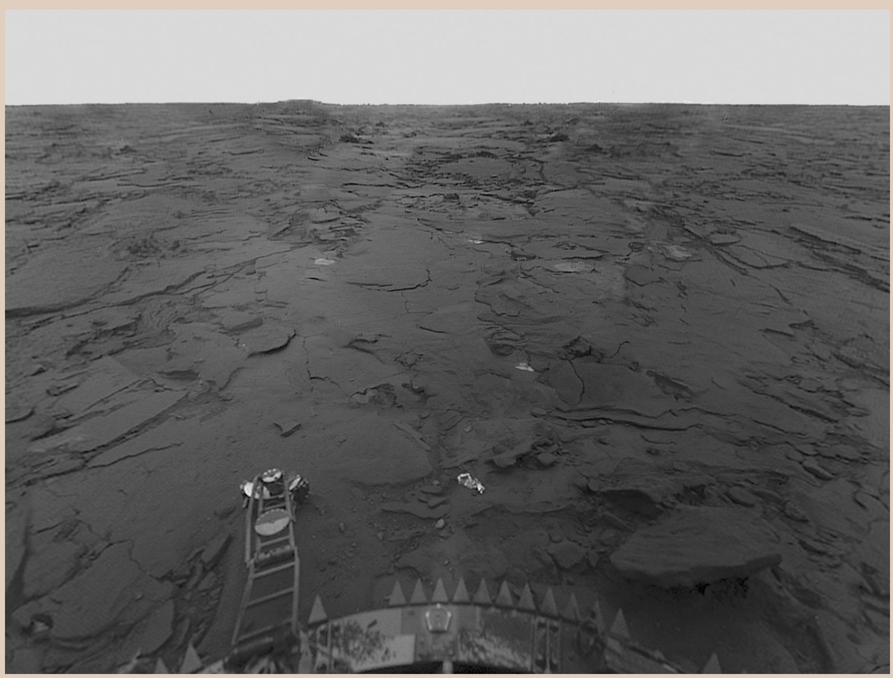
For us to truly understand the rules governing Earth-sized worlds in general, and our own Earth in particular, it's clear that we must return to Venus. We have a lot of catching up to do.

Lifting the Veil

Venus was actually a frequent target in the early days of space exploration, starting with NASA's Mariner 2 mission in 1962. Prior to then, Venus's distance to the Sun led some scientists—and many science fiction writers—to envision a lush, tropical world. Mariner 2 obliterated such visions when it returned measurements of torrid temperatures in the lower atmosphere. The numbers were so high that many researchers initially found them hard to believe: The surface temperature on Venus averages 465 degrees Celsius, hotter even than the planet Mercury. Mariner 2 also found that, unlike Earth, Venus has no protective, global magnetic field. Around the same time, radar measurements from a radio antenna at the Goldstone Deep Space Communications Complex in California showed that the rotation of Venus is *retrograde*: It spins in the opposite direction that it orbits the Sun, unlike all other planets except Uranus. Equally weird, it takes 243 Earth days for Venus to complete a single turn, by far the slowest of any planet in the Solar System.

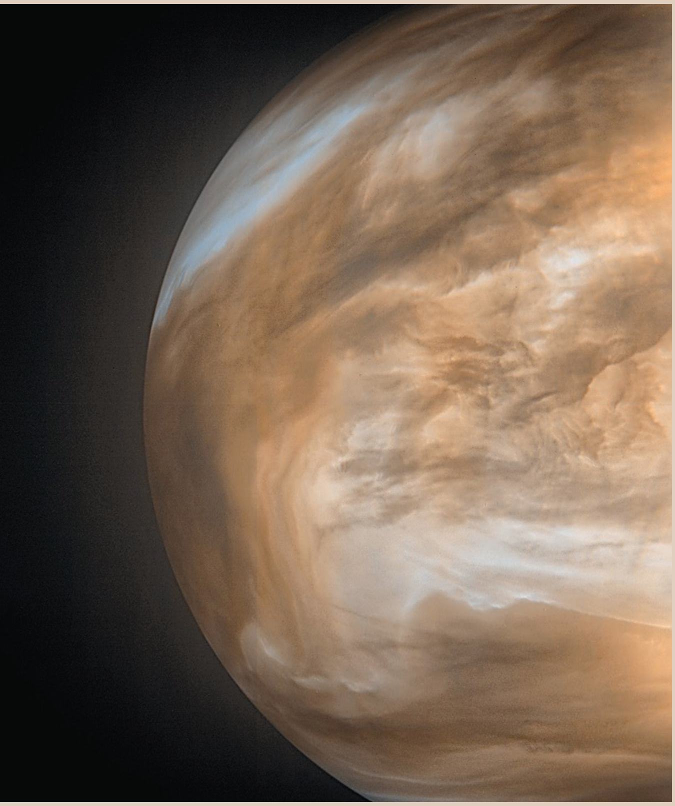
In 1967, the Soviet Union's Venera 4 probe plummeted into the Venusian atmosphere and took the first *in situ* measurements of its composition. The probe revealed that the atmosphere is 96.5 percent carbon dioxide, with negligible water vapor, and so thick that the surface pressure is 90 times that of Earth. The Soviet Union continued its focus on Venus, with Venera 8 in 1972 establishing that the planet's brilliant, silvery clouds are largely composed of corrosive sulfuric acid droplets.

Although Venus is bone dry today, apparently that was not always the case. Subsequent space missions measured the abundance of two forms of hydrogen—regular hydrogen and deuterium, which is chemically identical but twice as massive—and discovered

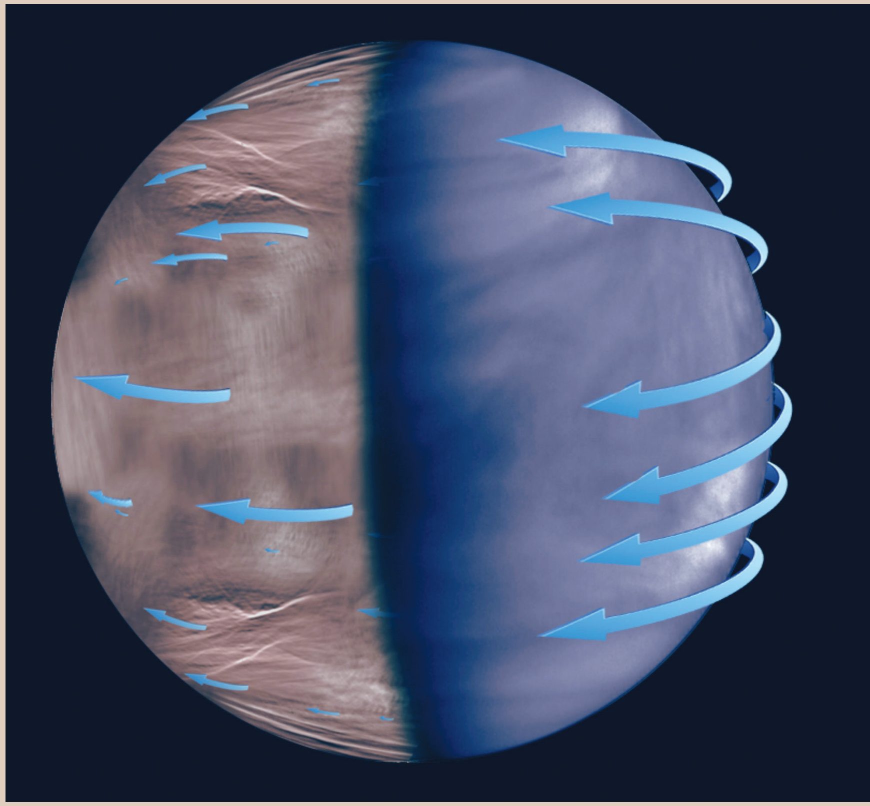


A rare look at the surface of Venus, captured on March 5, 1982, by the Venera 14 lander. The original scan (*below*) was reprojected (*right*) to yield a more natural perspective. Venus's dense atmosphere reduces sunshine to a dim, ruddy glow and causes distant objects to appear greenish. The flat plain surrounding Venera 14 is covered with rocks that resemble thinly bedded sediments. Chemical analyses suggest they might be lithified volcanic ash.

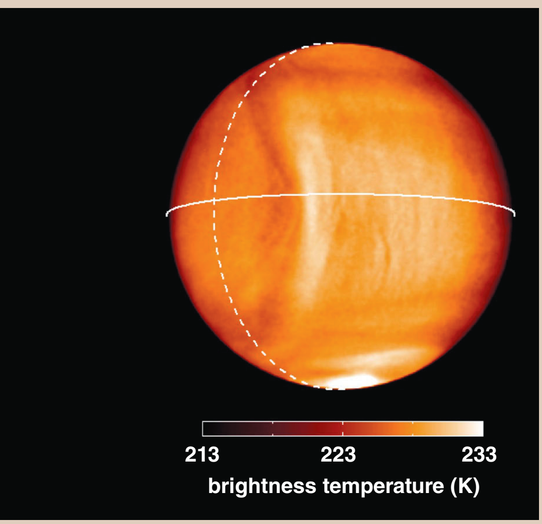
Courtesy of Donald P. Mitchell/Soviet space program



JAXA/DARTS/ISAS/Damia Bouic



Composite VIRTIS (images, left), AKATSUKI-UVI (image, right)/JAXA/ESA/J. Peralta, JAXA/R. Hueso, UPV/EHU.



Courtesy of ISAS/JAXA

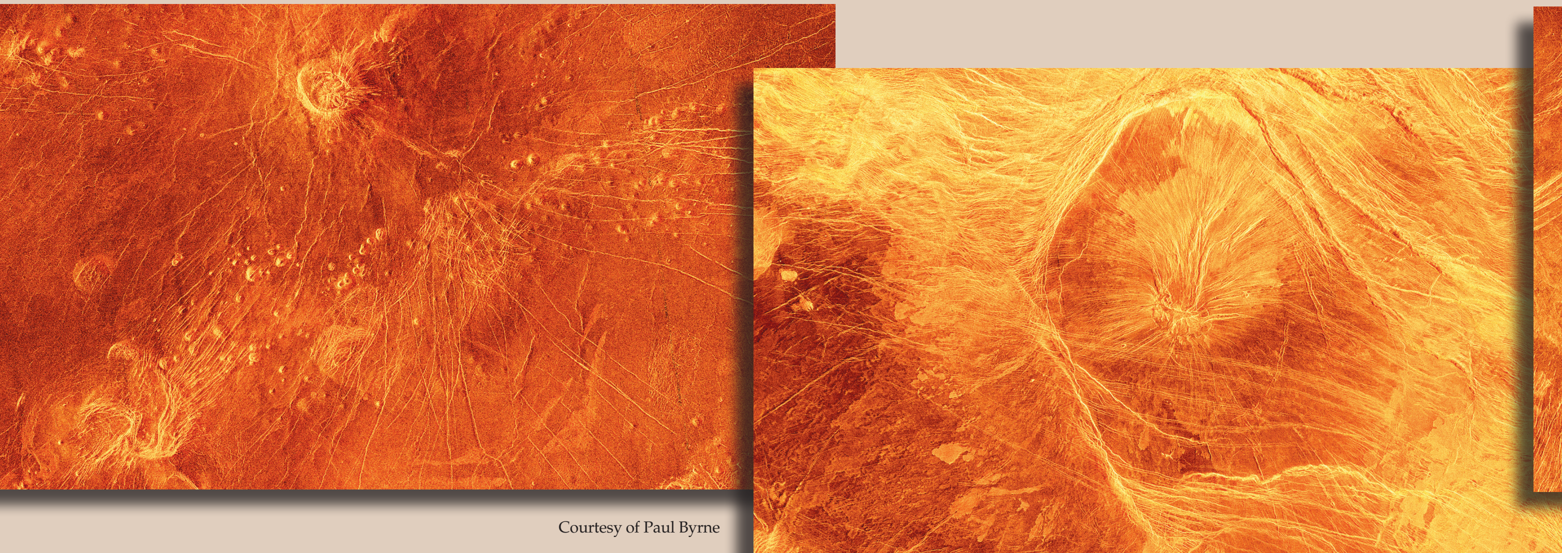
The Venusian atmosphere is like an ocean of carbon dioxide, trapping solar energy and producing complex flows. Viewed in thermal infrared by Japan's Akatsuki probe (*upper left*), the planet's nightside glows with heat. Clouds appear as dark silhouettes; changes in color correspond to variations in particle size or composition. Akatsuki also discovered a 10,000-kilometer-long standing wave (*left*) in the cloud tops. The upper atmosphere circles Venus every four days, 60 times as fast as the planet rotates. The Akatsuki and Venus Express probes studied this phenomenon, called *superrotation* (*upper right*).

that Venus's atmosphere contains about 100 times as much deuterium relative to hydrogen as Earth's does. Because Venus lacks a magnetic field, water molecules drifting into its upper atmosphere can be split apart into their constituent atoms and then stripped away by the solar wind. Deuterium, being more massive than regular hydrogen, is preferentially left behind by this process. The abundance of deuterium implies that the planet once possessed far more water than it does today.

NASA's Pioneer Venus mission, which deployed four atmospheric probes in December 1978, confirmed the intense temperatures and pressures measured by the earlier Soviet missions and reported back extreme variations in wind speed. The most complete information about Venus's atmosphere came from the final Soviet missions, Vega 1 and 2, in the mid-1980s. Both were headed to Halley's Comet, but along the way they flew past Venus, and each deployed a balloon probe and a lander. The balloons inflated and operated for more than a day, bobbing along at an altitude of about 55 kilometers, where the temperature and pressure are fairly close to those at sea level on Earth. During that time, the fierce high-altitude winds carried the balloons 11,000 kilometers, and their onboard accelerometers recorded dramatic up- and downdrafts. The main Vega spacecraft stopped relaying balloon data to Earth when they passed out of range, so nobody knows for certain how long the balloons operated before their batteries ran out.

The Vega 1 and 2 landers operated on the Venusian surface as well, but they were far from the first to do so. That honor belongs to the Soviet Venera 7 probe, which on December 15, 1970, touched down on Venus and returned a trickle of data despite being damaged on impact. Two years later, Venera 8 made the first fully successful touchdown, and broadcast a host of scientific measurements during its 50 minutes of life on the ground. All told, the Soviet Union landed 10 spacecraft on Venus, yielding the first images from the surface of another planet (Venera 9 in 1975), as well as the first off-world sound recording (Venera 13 in 1982). None of the landers returned data for more than two hours in the searing heat, but they were able to determine that the rocks on Venus are primarily basaltic, similar to much of the surfaces of Mercury, Mars, and the Moon, as well as the oceanic crust on Earth.

Because Venus's thick atmosphere and acidic clouds make it impossible to see the surface from above in visible light, NASA's Pioneer Venus orbiter (which arrived in 1978 and operated until 1992) and the Soviet Venera 15



Courtesy of Paul Byrne

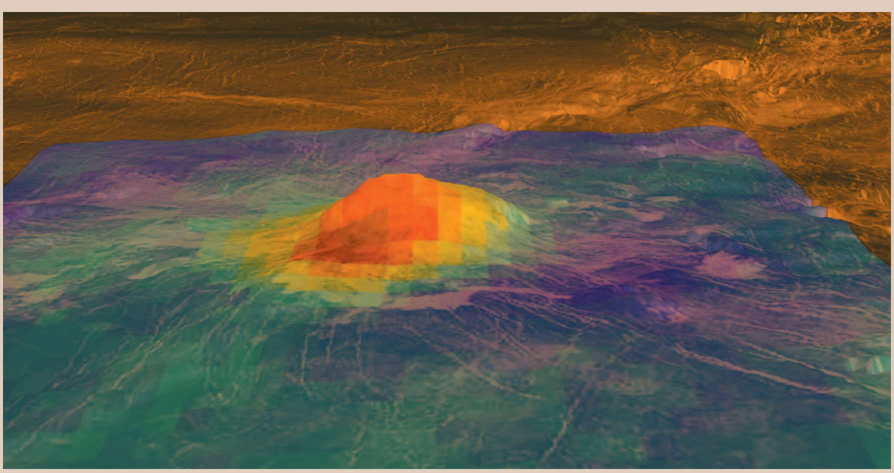
Radar views of Venus were captured by the Magellan probe and reprocessed by the author. Brighter, yellower colors correspond to surface material that is rough at radar wavelengths or that faces the incoming beam. Left to right: A chain of volcanoes, or "shield field"; Maram Corona, a distinctively Venusian structure, which may indicate where a huge rising plume of hot rock deformed the crust; Markham crater, an impact surrounded by a lopsided flow of melted rock or ejecta; and a portion of Ovda Regio, one of Venus's elevated tessera regions. The images are 370 kilometers, 700 kilometers, 550 kilometers, and 800 kilometers wide, respectively.

and 16 missions (which made orbit in 1983 and functioned for eight months) scanned the planet with radar. Those missions found evidence of extensive volcanic deposits and tectonic deformation, most notably in heavily fractured and folded *tesserae*, broad highlands that appear older than the lavas that surround them—although it was difficult to evaluate ages from those coarse images.

The issue of whether Venus's surface is ancient or young came into sharper focus with NASA's Magellan mission, which created what are still the most detailed radar maps of the planet over four years, starting in 1990. Magellan unveiled a world with diverse volcanic and tectonic features, including many bizarre and unfamiliar structures. Venus lacks the well-defined plate tectonics that characterize Earth. It also has no ancient terrains equivalent to the lunar highlands, Mercury's heavily cratered plains, or the bombarded southern uplands on Mars. Instead, it is broadly organized into a set of rift systems, low-lying plains, and the isolated highlands. In short, Venus looks utterly unlike Earth, but looks even less like any of the other rocky worlds of the inner Solar System.

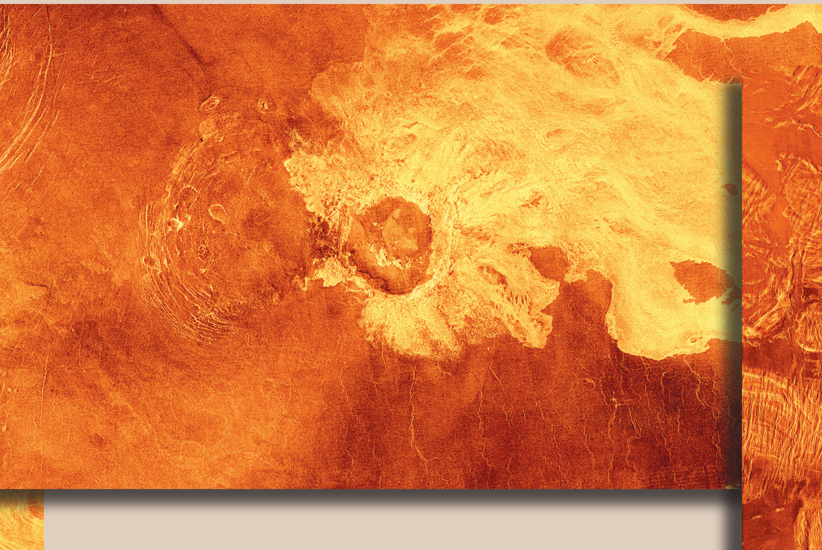
As a planetary geologist, I am fascinated by these unique, ambiguous landscapes. Some of my current work focuses on the tesserae, trying to deduce when and how they formed. It is possible they are so ancient that they predate the time when Venus entered its runaway greenhouse. If so, the tesserae might contain a record of the planet's earlier environment; for instance, those regions could contain sedimentary rocks and erosion features that formed in the presence of liquid water. Unfortunately, even the Magellan images do not contain enough detail to tell for sure. I've also studied the geology of the rifts, which may have jostled huge blocks of crust against one another. In a recent paper, a group of collaborators and I suggested that the jumbled Venusian crust might resemble what the Earth was like 3.5 billion years ago, when the modern process of plate tectonics was just beginning.

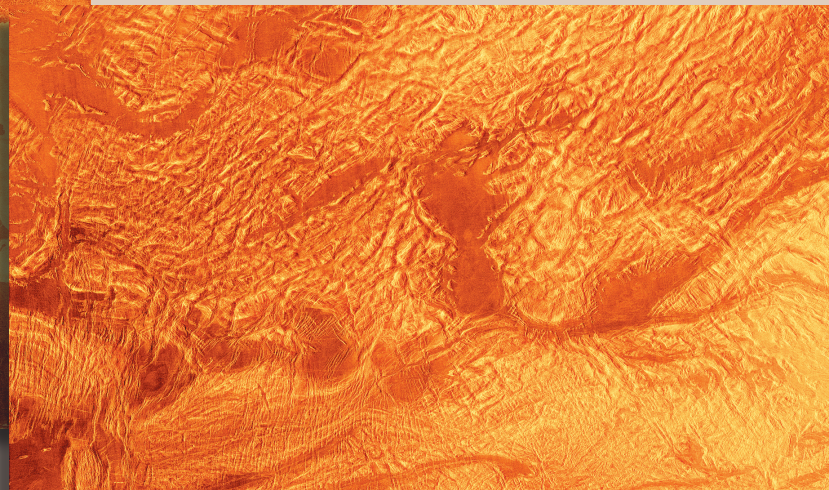
In the Magellan images, Venus looks like a planet that is still geologically active, but *how* active remains a matter of lively debate. Magellan's maps turned up a surprisingly small number of impact craters: Venus has fewer than 1,000 of them, and no gigantic impact basins equivalent to those on the Moon, Mercury, and Mars. Even more unexpected, those craters appear scattered randomly across Venus, with no region more heavily battered than any other. Extrapolating from the derived rate of impacts on the Moon, the surface of Venus is no more than 750 million years old on average. In contrast, most of Mars is more than 2.9 billion years old, and the surfaces of



NASA/JPL-Caltech/ESA

Young volcanism on Venus? The Venus Express orbiter measured the planet's surface emissivity, a measure of the way that materials emit infrared radiation. Several volcanic structures, including Idunn Mons (*shown above*), are higher in emissivity than rocks in the surrounding plains, implying that they are relatively fresh and unweathered. The Venus Express team concluded that those flows could be less than a few tens of thousands of years old, bolstering the case for modern volcanic activity on Venus—possibly even happening right now.





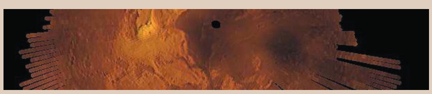
the Moon and Mercury have changed little in the past 3.8 billion years.

Planetary scientists have developed strongly divergent views about why Venus appears so relatively youthful. Some of my colleagues theorize that the planet was wiped clean by a global-scale catastrophe, with lavas pouring out all across the globe simultaneously. Others take a more nuanced view, suggesting that volcanism was episodic and localized, either taking place over a particular phase of Venus's life, or perhaps continuing throughout its history.

Modern Missions and Beyond

After Magellan, the exploration of Venus proceeded at a much slower pace. Russia's space-science budget collapsed after the breakup of the Soviet Union; meanwhile, the United States decided to concentrate its efforts on Mars. The European Space Agency (ESA) filled the gap with its Venus Express mission, which entered orbit about the planet in April 2006 and relayed data until late 2014. Venus Express collected long-term measurements of the atmosphere, including its thermal structure and its interaction with the solar wind, and recorded possible evidence of lightning. The spacecraft also measured how the surface of Venus emits infrared radiation. Those observations suggested that the lava flows in some regions might be quite fresh, geologically speaking—as little as 250,000 years old. Such plausibly recent flows strongly support the view that the planet remains volcanically active today.

At present, there is just a single spacecraft studying Venus: the Akatsuki probe from the Japanese Aerospace Exploration Agency (JAXA). Akatsuki entered orbit around Venus in late 2015, although it had originally been slated to do so five years earlier. An engine failure at the start of a critical maneuver required mission engineers to design a clever but circuitous new flight path. At the time of writing, Akatsuki is still operating at Venus, monitoring the planet's weather and characterizing the three-dimensional structure of the atmosphere. It has already made several notable discoveries, including a 10,000-kilometer-long wave centered above Aphrodite Terra, one of the planet's highlands. The probe has also shown that thermal tides might be the driving force behind a startling phenomenon at Venus called *superrotation*: The upper atmosphere of the planet rotates 60 times faster than the planet itself.



How can a world seemingly with the same starting conditions as Earth take such a vastly different path?

Venus will receive a smattering of additional attention from passing probes that use the planet's gravity either to speed up or to slow down in route to other planetary destinations. The joint ESA–JAXA BepiColombo mission to Mercury did a Venus flyby last October, and it will do another this October. NASA's Parker Solar Probe and several other upcoming missions will use Venus as a planetary slingshot, too. But these are limited, intermittent opportunities; there is no dedicated spacecraft slated to study Venus itself after Akatsuki.

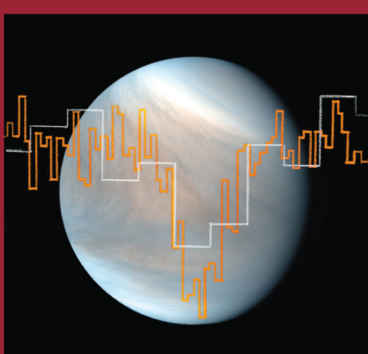
Fortunately, that situation may soon change. Several times per decade, NASA holds open competitions for its next planetary mission. NASA centers, research institutes, universities, and industry partners propose concepts that are then evaluated to determine their scientific value and novelty, technical readiness, and likelihood of success. For decades, missions to Venus have been passed over in favor of visits to Mars, asteroids, Jupiter, and Saturn's moon Titan. The odds look better this time around. In 2019, NASA received more than a dozen proposals for the Discovery program, which supports missions costing up to $500 million. When the agency chose four semifinalists for an additional nine-month study, two of the candidates were missions to Venus.

One contender is VERITAS (Venus Emissivity, Radio Science, InSAR, Topography, and Spectroscopy), managed by NASA's Jet Propulsion Laboratory. It would use radar to map Venus at 10 times the spatial resolution of Magellan, resolving details as small as 15 meters, and at up to 100 times the topographic resolution of the earlier mission. If selected, VERITAS would be able to settle the debate about when and how the surface of Venus was reshaped; explore whether the tesserae formed in the presence of liquid water; and detect active volcanism and other ongoing geological changes. Such data would transform our picture of Venus.

The other shortlisted NASA mission concept is DAVINCI+ (Deep

Life on Venus? *Really?*

Venus might seem utterly hostile to life, but it could have been habitable in the distant past. Even today, its cloudtops have temperatures and pressures much like those on Earth's surface. Given that similarity, a few researchers have wondered if strange dark material seen in its clouds, dubbed the "unknown absorber," could be hardy alien microbes. More compelling evidence of life on Venus arrived this past September, when an international team reported a radio-telescope detection of phosphine (shown as data lines over an ultraviolet image of the planet, *below*), a molecule that on Earth is produced mainly by bacteria. Other scientists soon questioned the validity of the result; the team stands by its findings. Even if the phosphine signal is real, it does not prove biological activity. But current life on Venus remains an open, albeit unlikely, possibility—one more reason to give the planet a much closer look.



Joanna Petkowska, https://www.joannapetkowska.com/

Atmosphere Venus Investigation of Noble gases, Chemistry, and Imaging, Plus), proposed by NASA's Goddard Space Flight Center. In addition to an orbiter equipped with infrared and ultraviolet cameras, DAVINCI+ includes a probe that would plunge down to the planet over one of its tesserae, taking images and measuring atmospheric composition all the way to the surface. The mix of trace gases in the atmosphere should determine whether Venus really did once possess far more water than it does today. High-resolution images from the DAVINCI+ probe during its descent, in turn, would make it possible to study the formation of Venus's highlands, and to look for the sedimentary rocks that would indicate the planet once had liquid water on the surface.

VERITAS and DAVINCI+ still face steep competition from two other Discovery-class proposals, one to visit Jupiter's volcanic moon Io, the other to fly past Neptune's giant, Pluto-like moon, Triton. NASA is expected to announce the winner (or possibly two winners) in the spring of 2021.

The U.S. space agency isn't the only one with its eyes on Venus. ESA is considering a "Medium-class" mission concept called EnVision, a radar-mapping orbiter that would also be equipped with ground-penetrating radar to scan the planet hundreds of meters beneath the surface; the agency plans to announce its decision in the middle of 2021. The Indian Space Research Organization has proposed Shukrayaan-1, a Venus orbiter that could launch in 2025. And Russia is developing the Venera-D concept, an updated incarnation of its storied Venera probes. The D stands for *dolgozhivushaya* (which means long-lasting), with the new lander designed to survive for a record three hours on the harsh Venusian surface.

Should any of these probes fly, there's likely to be a surge of public interest in the second planet from the Sun. We've seen that pattern repeatedly for Mars, as each new mission produces captivating results and a flurry of inspiring media coverage. For planetary scientists, the surge is already underway. Many of us—myself very much included—are hoping the 2020s will be "the Decade of Venus."

The Rosetta Planet

The breakthrough space voyages in the 1960s through the 1980s only began to put together the pieces of the enormous puzzle that is Venus—so much like Earth, yet so dramatically not. That stark disparity leaves scientists wondering which planet is the weirdo. Perhaps Venus suffered an early, out-of-the-blue catastrophe that slowed its rotation and weakened its magnetic field, leaving it defenseless against the Sun. Then again, maybe Venus is a typical midsize planet and Earth is the oddity, saved from a similar fate by an unlikely quirk of circumstance: the formation of the Moon, an unusually generous supply of water, or some other, as-yet-unknown factor. The best way to find out is to study what Venus was like in the distant past, and to determine how it became today's Hell Planet.

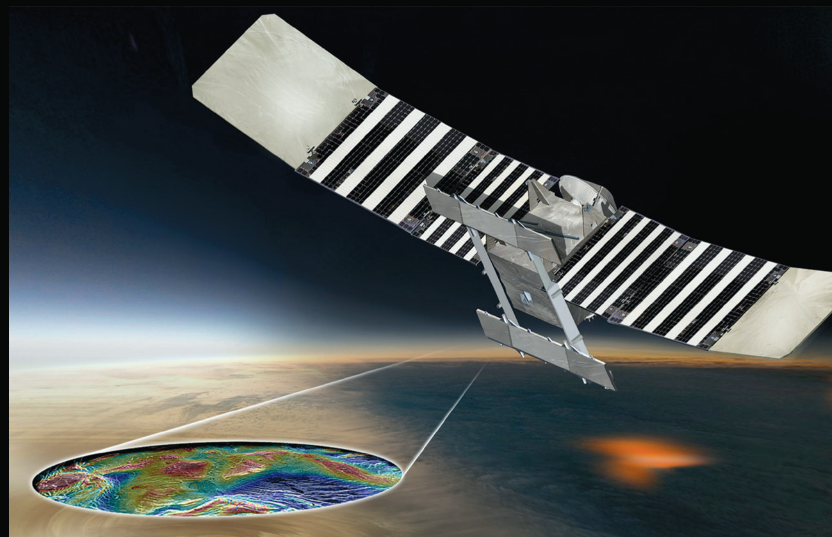
After the first space probes revealed Venus's true nature, some researchers began to theorize that the planet was not always so severe. It could have had oceans early in its history, when the young Sun was significantly less energetic and Venus received only about 30 percent more sunshine than Earth does today. In this scenario, Venus's surface was slowly but inexorably warmed by

It may be, then, that Venus's current state isn't the price of being relatively close to its host star, but simply bad luck.

the brightening Sun until its oceans started to evaporate too quickly to be replenished by rainfall. With a humid atmosphere trapping more and more heat, the planet's temperature continued to climb until the oceans boiled away. At that point, its fate was sealed.

Even if Venus started out wet and mild, that benign state might not have lasted long. Calculations by James Kasting of Pennsylvania State University in the 1980s indicated that Venus's runaway greenhouse kicked in early, in the first few hundred million years of its life. Earth's greater distance saved it from overheating, but only for a while. As the Sun continues to brighten, a similar fate could await our world about a billion years in the future.

Recent work has cast doubt on this interpretation, however. In a series of studies published since 2016, Michael Way of the NASA Goddard Institute for Space Studies and his colleagues argue that Venus's misfortune was not the fault of a changing Sun, but rather was self-inflicted. Way and his team theorize that the coincident occurrence of multiple enormous volcanic eruptions on Venus—each comparable to the one that resurfaced a huge swath of



Suzanne E. Smrekar/JPL/NASA

Courtesy of James B. Garvin/NASA

VERITAS (*top*) and DAVINCI+ (*left and above*) are being considered by NASA for flight to Venus in the late 2020s. VERITAS would make high-resolution radar maps and conduct spectral observations to determine the planet's material properties and geologic history. DAVINCI+ would drop through the atmosphere over the course of an hour, gathering the compositional information needed to reconstruct Venus's past climate.

northern Russia about 250 million years ago, triggering the Permian–Triassic mass extinction—was what wrecked the planet's climate. Such eruptions would have dumped so much carbon dioxide into the atmosphere over such a short period of time that the planet's rocks could not absorb it all. It may be, then, that Venus's current state isn't the inevitable result of being slightly closer to the Sun, but simply bad luck.

Way's climate models suggest that Venus could have remained habitable for up to three billion years, perhaps all the way until the time of a volcanic catastrophe. Given that simple microorganisms were present on Earth when the planet was about a billion years old (and possibly quite a bit earlier), some astrobiologists have speculated that life could have taken hold on ancient Venus as well—and even that Venusian microbes might survive to this day, floating in the planet's clouds. That last idea, long regarded as borderline fanciful, attracted considerable attention last autumn when a team led by astronomer Jane Greaves of Cardiff University in Wales reported a controversial detection of phosphine in the Venusian atmosphere. Phosphine is a rare molecule, commonly associated with anaerobic bacteria on Earth (*see box on page 36*).

And we can find out—if we return to Venus! Advances in spacecraft design, navigation, and instrumentation mean that a modern mapping effort would far surpass the results from Magellan. It could finally deliver the kind of high-resolution images my colleagues and I have been yearning for so that we can reconstruct the planet's geologic and climatic history in detail. An atmospheric probe like the one proposed for DAVINCI+ would vastly improve on the chemical measurements made by the Venus missions of the 1970s and 1980s. Even more exciting is the prospect of exploring Venus up close using next-generation landers or even rovers. Engineers at NASA's Glenn Research Center are developing electronics that could operate for weeks or months at Venus surface temperatures, addressing the most daunting technological issue that limited the lifetimes of the Soviet landers. Other experimental concepts would use clockwork-like mechanisms to replace electronics entirely, or rely on slow but potent winds to move a rover around using no onboard energy.

Studies of Venus also have implications that go far beyond our Solar System. Exoplanet searches have determined that our galaxy is full of warm, Earth-sized planets: A recent analysis of data from NASA's Kepler space telescope implies that there could be hundreds of millions of them. We don't yet have the observational capabilities to study the conditions on these strange new worlds, but the Second Planet from the Sun offers a natural, nearby laboratory for learning how Earth-sized planets form and evolve all across the universe. It can tell us why good planets go bad—and which one of these outcomes is the norm and which the aberration. If Venus is a rare misfire, then habitable, Earth-like worlds may be common around other stars.

The questions to investigate are profound. The timing is perfect, with tantalizing results from Akatsuki and planet-hunting telescopes begging for closer investigation. Four different space agencies have Venus mission concepts underway. We have the opportunity to begin a new era of exploration of the Second Planet. Will we take it?

(*References are available online.*)

Paul Byrne is an associate professor of planetary science in the Department of Marine, Earth, and Atmospheric Sciences at North Carolina State University. A geologist by training, he uses a combination of remote sensing and geospatial analytics, numerical and analogue modeling, and fieldwork to understand why planets look the way they do. He is a self-confessed Venus evangelist. Twitter: @ThePlanetaryGuy

New Twists in Earth's Radiation Belts

Rings of high-energy particles encircling our planet change more than researchers realized. Those variations could amplify damage from solar storms.

Daniel N. Baker

Late in the evening of January 31, 1958, a 32-ton Juno I rocket blasted into space from Cape Canaveral, Florida, lofting the Explorer I spacecraft into orbit. It was a mission of firsts: Explorer I was the first U.S. satellite (joining Sputnik 2, which had been launched the previous November by the Soviet Union). The satellite carried a pioneering scientific payload, prepared at the State University of Iowa by a team of researchers led by James A. Van Allen. And the instruments on Explorer I made the first revolutionary discovery of the Space Age: Earth is enshrouded in doughnut-shaped rings, or toroids, of high-energy, high-intensity radiation.

The discovery of those radiation belts—now called the Van Allen belts—revealed how Earth's magnetic field interacts with the space environment around it. The field, generated by Earth's molten metallic core and planetary spin, creates the magnetosphere, a magnetic bubble surrounding the planet; the size and shape of the magnetosphere change in response to the blowing of the solar wind, the constant stream of charged particles flowing from the Sun. The magnetosphere is crucial to life on Earth; it shields the atmosphere, as well as life on the surface, from damage by the solar wind and by even more energetic cosmic rays. But close in, Earth's magnetic field lines trap and accelerate free-floating particles, largely protons and electrons, and bounce them back and forth between the poles of the planet. Those zones of trapped, agitated particles make up the Van Allen belts that Explorer I flew through. It was discovered that the belts took the form of two concentric rings: The inner belt extends from an altitude of about 1,000 to 6,000 kilometers above Earth, whereas the outer belt spans from about 13,000 to 60,000 kilometers.

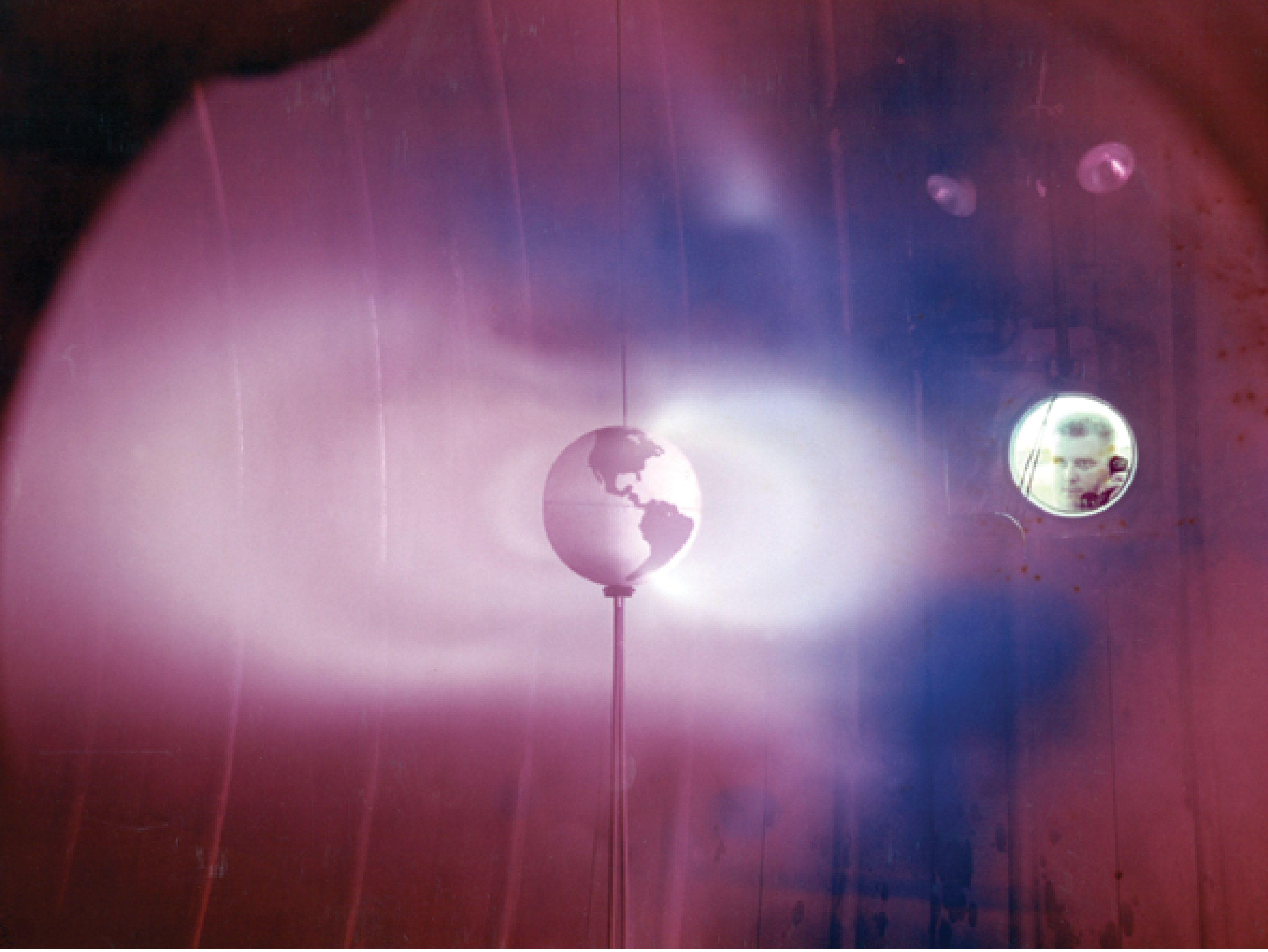
Earth's Van Allen belts are imperfect shields, however. High-speed particles can leak from the belts and collide with molecules in the atmosphere, giving rise to aurora displays. If there is a major magnetic eruption on the Sun, the resulting outrush of particles may break through the outer magnetosphere and overload the Van Allen belts in more destructive ways. The rapid injection of particles into the belts can damage the circuitry and solar panels on satellites in orbit; swarms of protons and electrons released when solar wind particles crash into the atmosphere induce electrical currents that can overload terrestrial power systems and cause blackouts.

Almost exactly a century preceding the Explorer I launch, on the night of August 28 to 29, 1859, people around the world got to witness what happens when an enormous solar storm overwhelms Earth's magnetosphere. *The New York Times* reported that thousands of New Yorkers watched "the heavens…arrayed in a drapery more gorgeous than they have been for years." An even more spectacular aurora display occurred on September 2, when the sky lit up as far south as Central America in the Northern Hemisphere. Disturbances in Earth's magnetic field were so powerful that magnetometer readings were driven off their scales. Telegraph networks were unusable for nearly eight hours in most parts of the world due to high-energy particles in the atmosphere. In several regions, operators reported that their telegraphs were sparking from the electrical current induced by the aurora. Earth had experienced a one-two punch of solar storms the likes of which have not been recorded since.

Humanity was just beginning to develop electrical technology in 1859. There were no high-power electrical lines crisscrossing the continents, nor were there sensitive satellites orbiting Earth. In 1989, just before the rise of the Internet and GPS systems, a smaller but still potent solar storm demonstrated the heightened risk. The 1989 storm induced huge ground currents that knocked out Quebec's electrical power grid and caused problems at 200 sites in the United States, particularly in regions situated on igneous rock because it resists conduction and therefore flows current into nearby wires. If another solar event like the one in 1989 happened today it could disrupt global communications, causing chaos for days. Another 1859-style superstorm could knock out some power grids and communications networks for weeks or more.

Our Sun operates on an 11-year cycle of activity, and today it is near the maximum of that pattern, meaning it could at any time produce large-scale events. In mid-July 2012, a solar storm of immense power narrowly missed the Earth; had it happened a week earlier, the planet might have been in the direct path of

Daniel N. Baker is Director of the Laboratory for Atmospheric and Space Physics, as well as a professor in atmospheric and planetary sciences, and in physics, at the University of Colorado-Boulder. He obtained his PhD in 1974 with James A. Van Allen at the University of Iowa. He is currently an investigator on several NASA space missions, including the MESSENGER mission to Mercury, the Magnetospheric MultiScale (MMS) mission, and the Radiation Belt Storm Probes mission (now called the Van Allen Probes mission). E-mail: Dan.Baker@lasp.colorado.edu



In this 1966 photograph, a plasma thruster at NASA's Lewis Research Center (now John H. Glenn Research Center) simulates the Van Allen belts, rings of radiation that surround the Earth. The rings had been discovered just a few years before, and a new mission shows there is still much to learn about the impact they can have on the Earth and its satellites. (Image courtesy of NASA.)

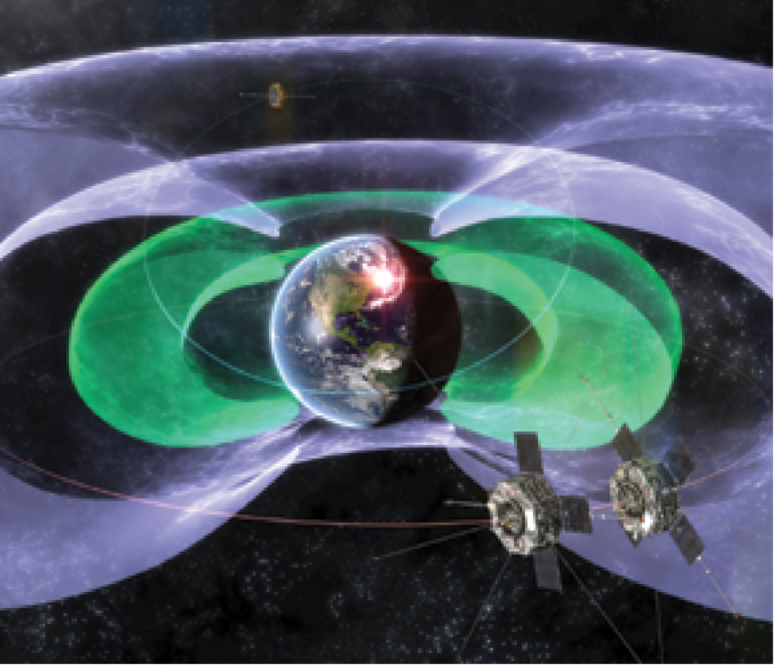
the blast. My colleagues and I are vigorously pursuing studies of space storms and the changes in our near-Earth space environment, which we lump under the term *space weather.* There is a pressing need for our technological society to understand in ever better detail the workings of the space environment around us. A clearer picture of the dynamics of the Van Allen belts is one important piece of this puzzle.

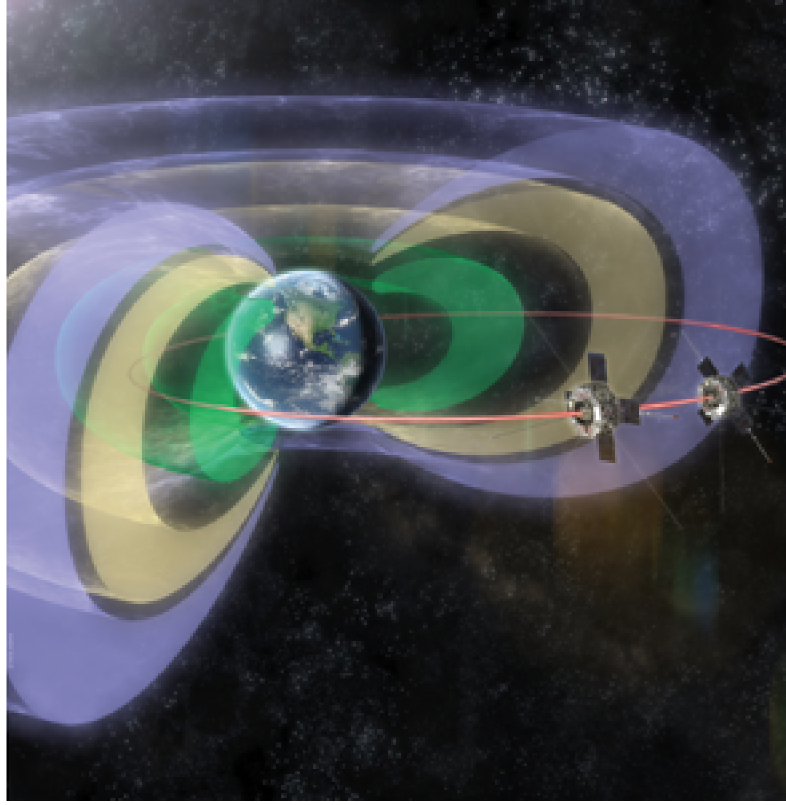
Space Storm Damage

What happens to satellites during space storms is of great practical importance. After the pioneering work of Van Allen and his coworkers in the United States, along with their counterparts in the Soviet Union, there was an explosion of interest in the use of space for human needs. Over just a few years in the late 1950s and early 1960s, space hardware went from technological demonstration and scientific curiosity to full-fledged societal imperatives. Earth satellites were launched into space to meet needs for communication, navigation, weather observations, remote Earth sensing, and military reconnaissance. Today the Earth is circled by spacecraft from just above our atmosphere to distances of tens of thousands of kilometers above Earth's surface. It would be almost inconceivable to try to imagine our modern U.S. society without the capabilities provided by spacecraft systems. But any of the many hundreds of spacecraft operating in Earth orbits today can be damaged by space radiation if the circumstances are right. In 2003, 46 of the 70 satellite failures reported that year occurred during a geomagnetic storm in October.

When high-energy protons and other ions hit orbiting spacecraft, they often leave ionization tracks in electronic chips. These tracks can upset spacecraft computer memories and otherwise disrupt sensitive electronics. As a result, satellite solar power panels may be damaged, optical tracker systems may become confused, and spacecraft command-and-control software may be scrambled. High-energy protons and ions may also injure, and potentially kill, astronauts who are in space during a major solar particle event. Manned launches have had to be rescheduled as a result, a major obstacle to long missions such as ones that might go to Mars. The high-energy protons in the inner Van Allen zone are especially a continuing risk to satellites and humans alike.

Energetic electrons in the space environment can also be devastating to spacecraft. They can readily penetrate even thick spacecraft shielding and bury themselves in insulating materi-





Before the 2012 launch of what is now known as the Van Allen Probes mission, the Van Allen belts were believed to be two concentric rings of particles (*left*). New data have shown that a third temporary ring (*right, yellow*) can form due to solar storm activity. (Illustrations courtesy of Dr. Andy Kale, University of Alberta.)

als, such as coaxial cables or electronics boards, deep within spacecraft systems. As charge builds up in the insulating materials, a powerful internal electrical discharge can occur, much like a miniature lightning strike. Numerous recent spacecraft failures have been attributed to this mechanism.

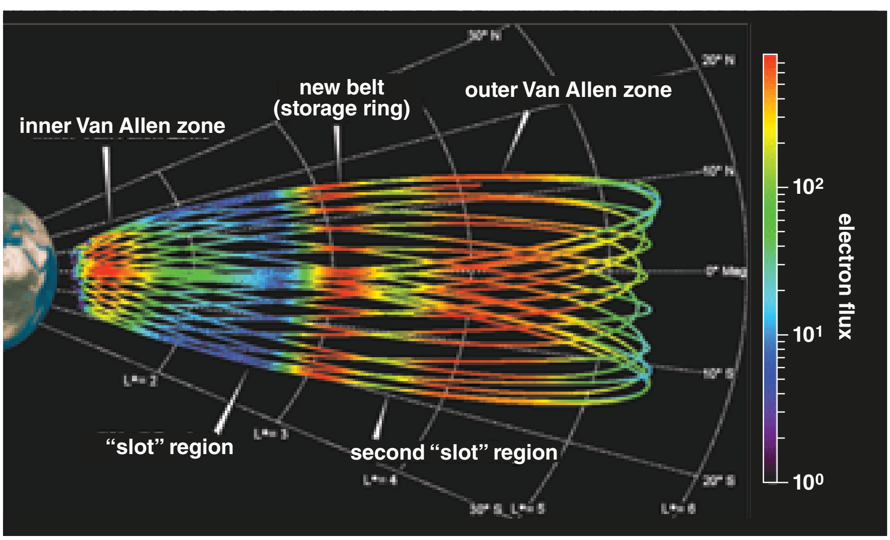
Another space weather effect is known as surface charging. Lower energy electrons cannot penetrate the shielding but can accumulate on insulating satellite surfaces. As with interior insulators, charge buildup on the surface may lead to a powerful, disruptive discharge, generating electrical signals in the spacecraft's vicinity that can scramble and disorient the satellite and its subsystems.

A Third Belt

In light of the world's dependence on Earth-orbiting platforms, it must be realized that every one of these spacecraft fly through—essentially continuously—the high-energy radiation environment that Van Allen's group discovered over five decades ago. Thus, one of the most enduring and persistent aspects of space weather is the hostile radiation belts girding the Earth. Probes have returned data showing that the Van Allen belts wax and wane in intensity, depending on both local conditions and Sun activity. Even 50-plus years after their discovery, we still need a deeper and more insightful comprehension of the Van Allen belts' behavior.

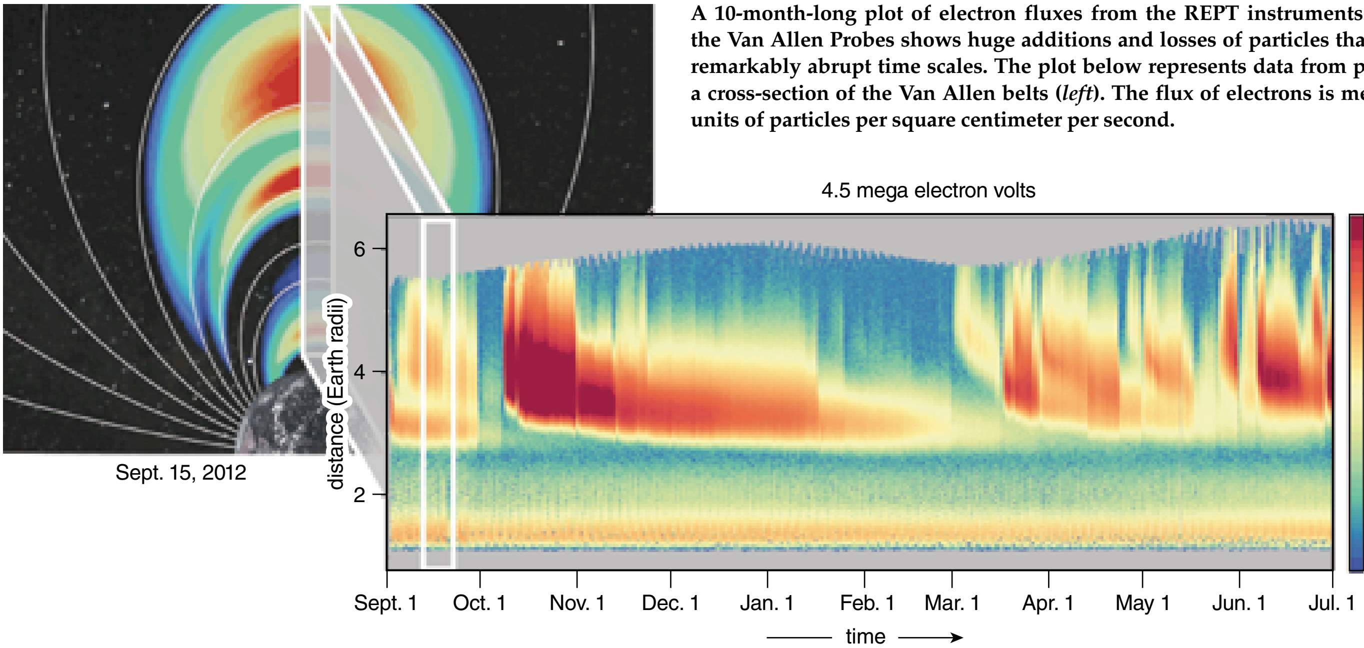
About a decade ago, NASA began developing a program called Living With a Star. This program name acknowledges that we on Earth live in the outer atmosphere of a magnetically active star—our Sun—that exerts a huge influence over all aspects of the environment on our planet. We rely on the Sun for heat and light, of course, but we also must endure the fits of temper that it exhibits when it releases huge x-ray emissions in a powerful solar flare, or when it erupts 10 billion tons of hot plasma at several million miles per hour in dramatic expulsions we call *coronal mass ejections.* These immense solar storms can, in turn, pump up the Van Allen belts to an extraordinary degree, making the radiation zones around Earth immensely more dangerous for days or even weeks on end.

The Living With a Star spacecraft program got under way on February 11, 2010, with the launch of the Solar



In their first three days in orbit, the Van Allen probes showed the formation of a new ring in Earth's radiation belts. Data were captured by an instrument on the probes called the Relativistic Electron-Proton Telescope (REPT). The spatial distribution of the high-energy electrons in the belts is overlaid on the orbital trajectories of the spacecraft. (Unless otherwise indicated, images are courtesy of the author.)

A 10-month-long plot of electron fluxes from the REPT instruments on board the Van Allen Probes shows huge additions and losses of particles that occur on remarkably abrupt time scales. The plot below represents data from particles in a cross-section of the Van Allen belts (*left*). The flux of electrons is measured in units of particles per square centimeter per second.



Dynamics Observatory. This spacecraft has been returning huge volumes of data about the Sun and the outer solar atmosphere called the corona, as well as images of the Sun that are truly breathtaking in their beauty and detail (see *Sightings*, July–August 2014).

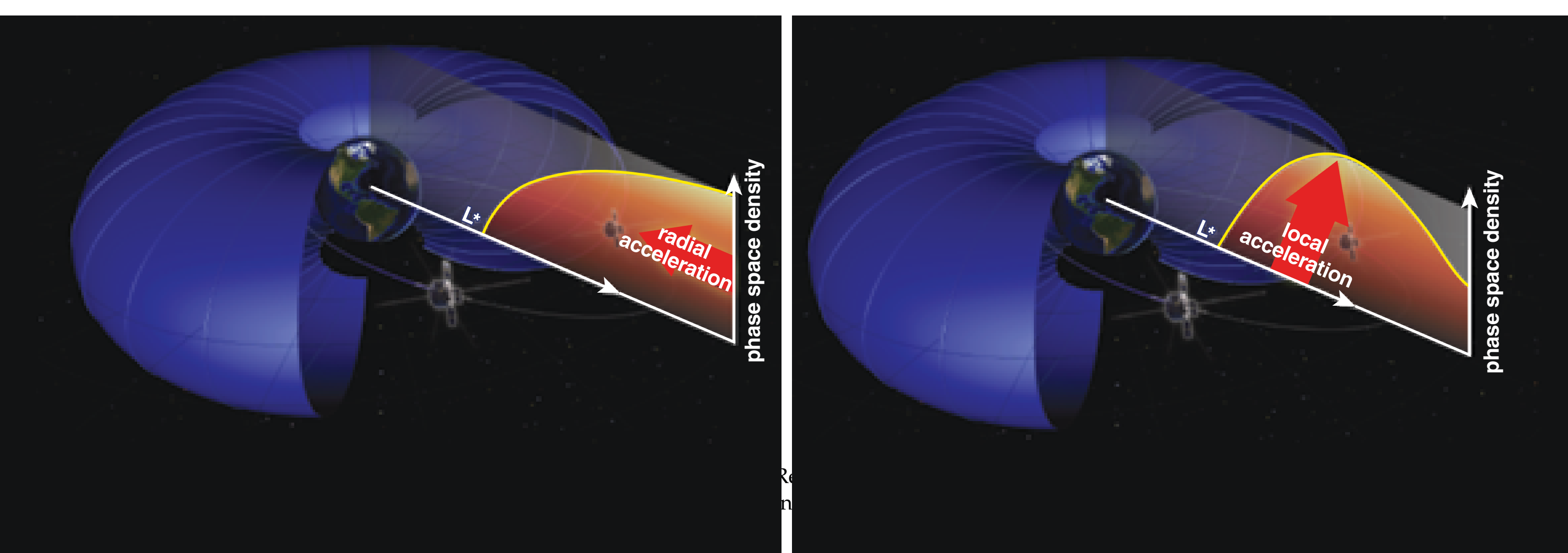
The second wave of the program, dubbed the Radiation Belt Storm Probes project and later renamed the Van Allen Probes, was geared toward studying the Van Allen radiation belts in all their complexity and variability. The aim of this mission was to launch two identical Earth-orbiting spacecraft that fly in large elliptical, equatorial orbits that would carry them all the way through both the inner and outer Van Allen belts. With comprehensive measurement capabilities in each of the twin spacecraft, the multiyear mission would examine the Van Allen belt regions at a level of detail never before seen. The satellites were launched on August 30, 2012.

After most NASA spacecraft launches, experiment teams wait patiently for several months as research instruments on board are turned on one at a time, slowly ramped up to full power, and generally tested to make sure they work at full capacity. That was the plan for the Van Allen Probes as well. Our research team, however, urged that our instrument, the Relativistic Electron-Proton Telescope (REPT), be turned on just three days after launch. The reason for our haste was that another Sun-monitoring satellite called SAMPEX (Solar, Anomalous, and Magnetospheric Particle Explorer) was about to re-enter Earth's atmosphere and be destroyed. We wanted the instruments on the two missions to overlap so that we would have continuous, calibrated data. NASA agreed to our request.

It was a lucky decision. On August 31, just before REPT turned on, a long filament of solar plasma erupted out into space, sending a powerful solar coronal disturbance toward Earth. When that storm started to reach Earth a day later it hit the radiation belts and caused them to change dramatically. To the great satisfaction of our team, the REPT instruments worked well from the moment they were turned on, on September 1. REPT observed freshly accelerated particles trapped in the belts, and recorded their high energies as the belts increased in size and strength.

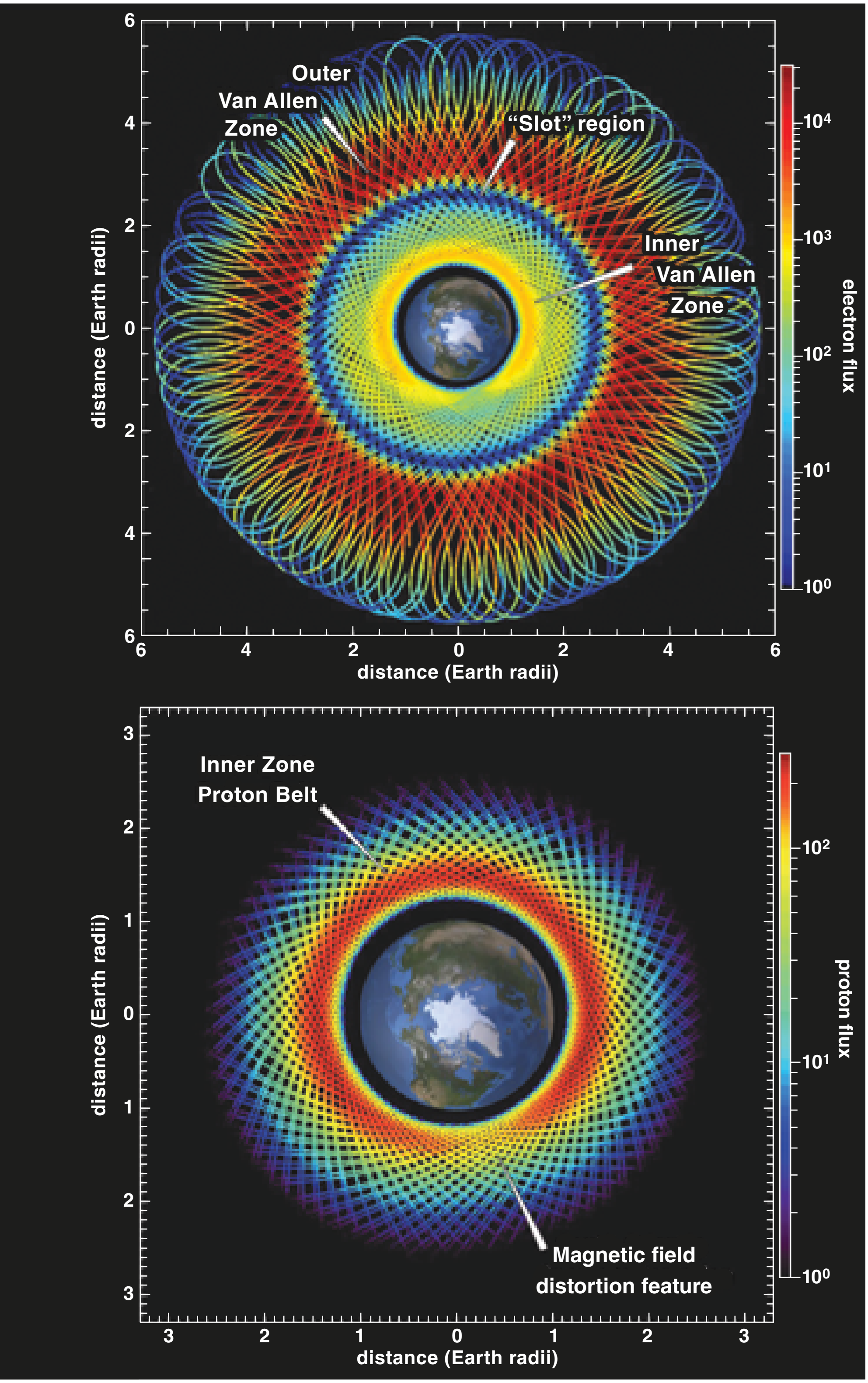
Then something happened that no one had ever reported before: The particles settled into a new configuration,

Two different theories show how particles in the Van Allen belts could be accelerated: Particles may slowly diffuse inward from the magnetosphere (*left*) or they may undergo intense local acceleration within the heart of the outer Van Allen zone (*right*). (Illustration courtesy of NASA/ G. Reeves/M. Henderson.)



showing an extra, third belt between the two known belts. Within mere days of launch, the RBSP spacecraft showed us something that would require rewriting textbooks about how the Van Allen belts can be configured. The third belt was dubbed the storage ring in our publications.

Our team uses color coding of particle intensities to demonstrate how the radiation belts changed in time and space over the first full year of operation. As shown on page 377, this long-term summary features fascinating examples of abrupt onset of electron acceleration, long periods of gradual inward radial transport of energetic particles, and striking examples of abrupt electron losses. The three-belt radiation zone structure from September 2012 is visible, as is the return of this remarkable feature in March 2013.



Belt Drivers

In their first two years in orbit, the instruments on the Van Allen Probes have worked exceptionally well, and our science teams are excited about a flood of observations coming to us with unprecedented clarity and quality. This is the first time we have been able to gather such a complete set of data about the belts, with the added bonus of watching from two separate spacecraft that can better show how events sweep across the area in time. Spotting something new in space, such as the third radiation belt, has more implications than the simple knowledge that such a feature is possible—for instance, it shows that extremely high-energy particles can appear and disappear almost in the blink of an eye. In a region of space that remains so mysterious, any observations that link certain causes to specific effects adds another crucial piece of information to the long-standing puzzle of what drives these belts.

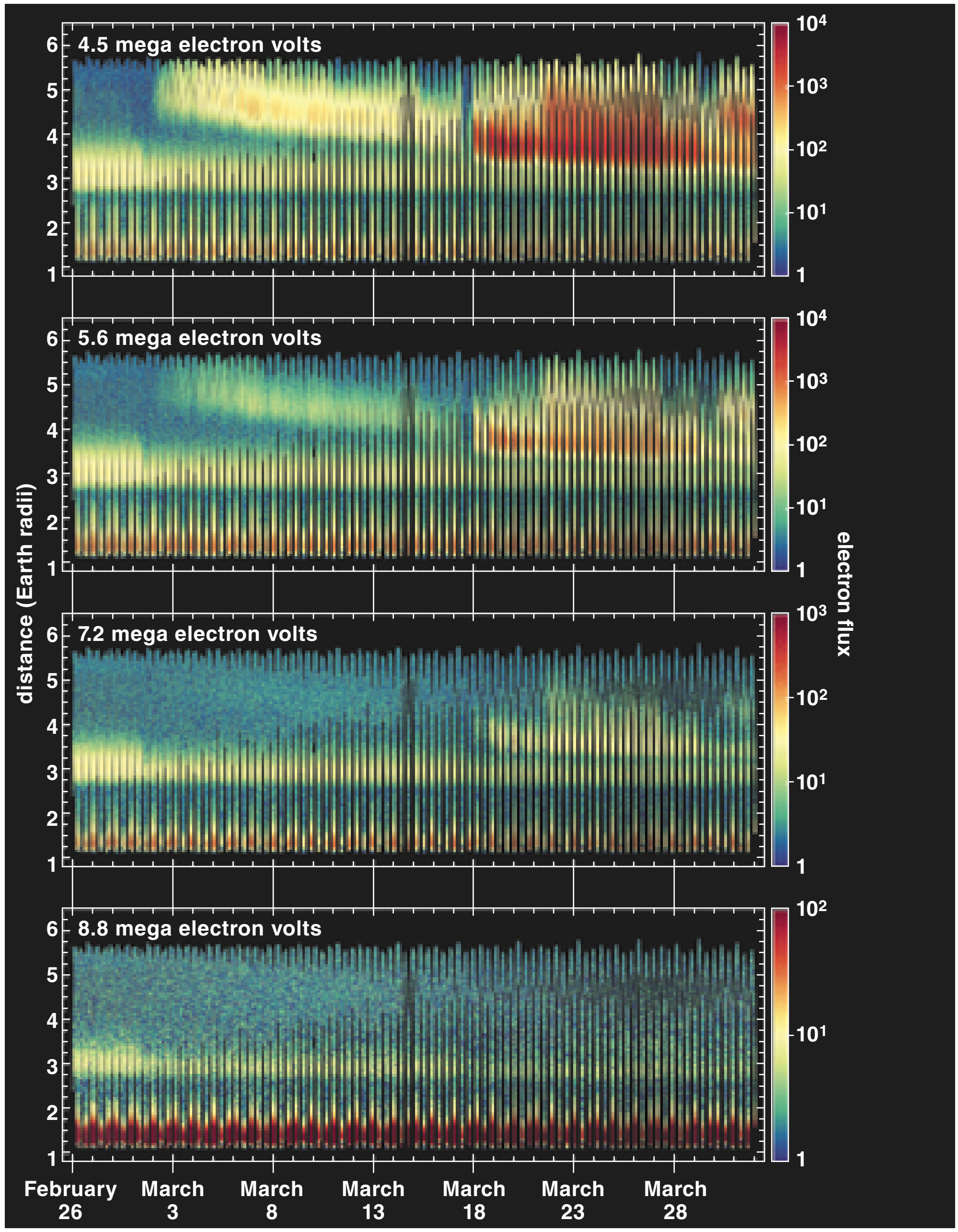
The probes have also filled in details about the inner workings of the Van Allen belts that had previously been merely speculative. For several years after the 1958 discoveries of Van Allen and his coworkers, researchers theorized that the radiation belt electrons come from the distant reaches of Earth's magnetosphere. Theorists proposed that as the particles drifted closer to Earth and encountered stronger magnetic fields, electrons would be accelerated and would form into a ringlike configuration. But this type of acceleration process would take days to weeks and best describes radiation belts that vary only gradually over time.

In the 1990s, satellites such as SAMPEX began to reveal that the energy and density of the Van Allen belts changed much more quickly. Consequently, a competing theory for the origin of the

Electron fluxes in the outer Van Allen belt (*top*) and proton fluxes in the inner belt (*bottom*) are projected in a "spirograph" pattern over the northern pole of the Earth. The projections show the bands of particle energies within each belt. The gap shown in the inner belt at the magnetic field distortion feature is a result of the tipped magnetic pole of the Earth.

belts' electrons began to take hold: It was suggested that charged particles do not come from far out in the system, but are produced locally within the heart of the outer belt. This would occur when electric fields within the belts take low-energy electrons, which are ever present in all regions of space, and accelerate them to nearly the speed of light. This process could alter the density and energy of the belts on scales of seconds to hours, a theory that matched better with the observations from the 1990s.

Unfortunately, older satellite observations were too sparse in spatial coverage and the spacecraft were not designed to measure rapidly changing properties of the belts at different locations, as would be needed to fully distinguish between the two proposed acceleration mechanisms. Those data shortcomings changed drastically with the launch of the Van Allen Probes. In early October 2012, a week after a solar storm had swept the outermost Van Allen belt of virtually all of its high-energy electrons, the two probes recorded a nearly 1,000-fold jump in electron density in less than 12 hours (*as shown in the figure on page 377*). The twin set of spacecraft observations clinched the case for electric fields deep within the belt accelerating the electrons. But research my colleagues and I published in early 2014 showed that the outer radiation belt developed at different time scales: Particles would be lost from the belt by a gradual inward diffusion, over a period of months, and then would be replenished rapidly by strong solar wind events. Therefore several mechanisms of electron energization are in play in the belts, often acting almost concurrently or at least in rapid succession.

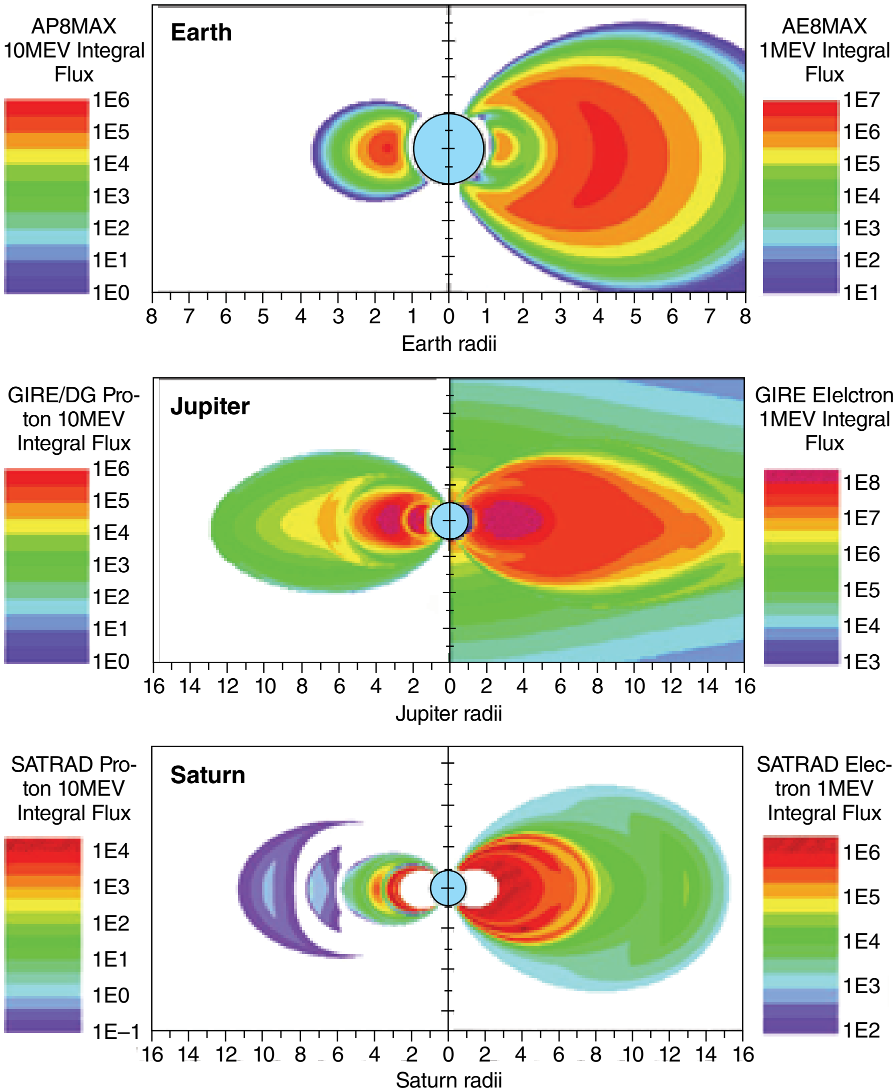


During March 2013, a solar storm again induced the formation of the ghostly third belt or storage ring, as shown by four channels from the REPT. The storage ring remained for only a matter of weeks before dissipating again.

We have, with the Van Allen Probes October 2012 event, been able to readily distinguish between the different ways that particles can be accelerated. In this way, the Van Allen Probes have seen right to the heart of the magnetospheric acceleration process. The measurements clearly show that substantial particle acceleration can occur locally in the core of the outer radiation zone. This work shows that the frequency of some of the electromagnetic waves in the belt matches the frequency at which electrons travel around the local magnetic field, a synchrony that makes it easy to pump up the charged particles.

My colleagues at the University of California at Berkeley, led by Forrest Mozer, report in a July 2014 paper that their analysis of data from the Van Allen Probes, as well as computer simulations they created, show that a twofold process likely governs the acceleration of particles to high speeds in the Van Allen belts. Initially, very short duration pulses of electric field in the belts can accelerate the particles somewhat. After that, the particles are of the correct energy to be in resonance with a type of electromagnetic wave called a *whistler* that is generated by lightning discharges in the atmosphere. Several interactions with the whistler waves can kick the particles up to great energy levels. My team and I had published an analysis in 2013 proving the definitive involvement of whistler waves in particle acceleration.

It is challenging to describe measurements of the Van Allen belts in a graphically accessible way, so we are continually experimenting with new data visualizations (*see page 378*). One of our newest images shows the paths of the two Van Allen spacecraft as they move through the different zones of radiation. The satellites precess in elliptical orbits around the Earth, forming



The planets within our Solar System, such as Jupiter and Saturn, produce their own radiation belts of distinctive sizes and shapes, and improved knowledge of Earth's radiation belts may also help in understanding such structures around other worlds. (Planets are not shown to scale.) (Adapted from an illustration by Henry Garrett, JPL.)

a spiral pattern as seen from a fixed, Earth-centered coordinate system. Their data, measuring energetic electron and proton fluxes, are projected onto the geographical equatorial plane. The resulting orbital pattern looks much like a drawing created by a spirograph children's toy, showing in the colored bands the weak inner Van Allen zone population of electrons, the slot region between belts, and the relatively stable (at least during this quiet six-week interval) outer Van Allen zone. Detailed features can be discerned in this display such as the relative paucity of inner zone energetic electron fluxes over the American sector; we believe this corresponds to the effect related to the Earth's offset, tilted magnetic dipole.

Examining the color-coded flux profiles reveals that for the period shown (November 2 to December 15, 2012) the energetic electrons had the familiar, clear two-belt structure. The very high-energy protons, on the other hand, are seen only in the inner Van Allen zone and are confined to the region quite close to Earth. The proton data in this format clearly show the inner belt region and also show longitudinal asymmetries in particle fluxes associated with the Earth's offset, tilted dipole magnetic field. This leads to an apparent reduction of particle fluxes over the American sector due to distortions in the Earth's magnetic field.

One Illuminating Month

The first two years of Van Allen Probes observations have demonstrated the immense benefits of near-equatorial dual-spacecraft measurements of highly relativistic electrons. Obtaining excellent temporal, spatial, and energy resolution has been key. When the solar wind is relatively calm (at speeds of about 300 to 600 kilometers per second) the outer radiation belt is characterized by slow, steady diffusive transport and gradual particle loss. However, the appearance of enhanced solar wind forcing, from transient coronal mass ejection drivers or high-speed solar wind (with velocities greater than 500 kilometers per second), quickly produced strong increases in particle energies and densities.

The Van Allen Probes observations have illuminated the intimate relationship between external solar wind forcing and internal magnetospheric particle acceleration and transport by both solar wind streams and coronal mass ejection–driven geomagnetic storms. Data collected in March 2013 covered a particularly important interval to see how particle acceleration in the Earth's radiation belt works (*see page 379*). An event early in the month, associated with a high-speed solar wind stream, with a velocity greater than 650 kilometers per second, gave rise to a strong electron acceleration event that was centered at relatively large distances of five to six Earth radii, or about 35,000 kilometers. The population formed after the passage of the leading edge of the fast solar wind stream. For the next two weeks, our team observed that the core of the high-energy electron population diffused inward in radial distance, all the while exhibiting gradual diminution of magnetic fluxes, suggesting continual weak losses of particles.

The winding down of this long, diffusive acceleration event came to an abrupt end on March 17, 2013, when a strong interplanetary shock wave impacted Earth's magnetosphere. As the shock passed, the solar wind speed jumped from about 425 to about 725 kilometers per second, and the solar wind magnetic field also showed very large changes. This event led first to extensive radiation belt depletion, but then to almost as rapid a reappearance of a seemingly new radiation belt population very deep in the outer belt. This new population was reminiscent of the October 2012 event.

Concurrent observations from the electric field and magnetometer experiments on board the Van Allen Probes spacecraft show that for both of these March 2013 events, the magnetospheric boundaries were forced deeply inward toward the Earth. Hence, we see that relativistic electron acceleration occurs when (and perhaps only when) the outer Van Allen zone is situated well outside its normal magnetospheric location. Only then can electromagnetic waves interact with much lower energy "seed" particles that are necessary for relativistic electron production. Our team is still exploring in greater depth the details of this acceleration process for March 2013 as it relates to the magnetospheric configuration and wave properties.

Belts on Other Worlds

Even as the Van Allen Probes are allowing us to explore in detail the highly efficient particle accelerator operating around our planet, it also is leading to a better understanding of the equivalent processes taking place around other worlds. We can use the Earth's radiation belt system as our most accessible local laboratory to study particle acceleration and transport mechanisms. We can then extend these lessons to the magnetospheres of other planets with strong magnetic fields, including Jupiter, Saturn, and Mercury (*see figure on page 380*). In addition, we can now also extend the lessons of the Van Allen Probes to the thousands of planets being discovered around other stars. We can even apply these lessons to the extremely powerful magnetospheres of neutron stars—the collapsed remnants of supernova explosions—and, perhaps, to the magnetic fields that confine streamers of plasma on vastly larger, galactic scales.

In the 55 years since the launch of Explorer I, space scientists have made tremendous advances in their understanding of the magnetic and particle environment around the Earth. It is a fitting legacy to Van Allen and his pioneering discoveries that the Van Allen Probes continue to reveal new phenomena about this energetic and perilous region of space. Data from the Van Allen Probes, combined with readings from other satellites, are already being used to improve computer models that can incorporate real-time data of space activity and help forecast what's happening in near-Earth space. When we look upon physical systems with new eyes—even those systems we think we know so well—we can make remarkable discoveries and gain amazing new insights. This possibility is a marvelous aspect of the endless frontier of space research.

The two Van Allen Probes recorded a nearly 1,000-fold jump in electron density in less than 12 hours.

Bibliography

Baker, D. N., et al. 2014. Gradual diffusion and punctuated phase space density enhancements of highly relativistic electrons: Van Allen Probes observations. *Geophysical Research Letters* 41:1351–1358.

Baker, D. N., et al. 2013. James Van Allen and his namesake NASA mission. *Eos, Transactions of the American Geophysical Union* 94:469–470.

Baker, D. N., et al. 2013. The Relativistic Electron-Proton Telescope (REPT) instrument on board the Radiation Belt Storm Probes (RBSP) spacecraft: Characterization of Earth's radiation belt high-energy particle populations. *Space Science Reviews* 179:337–381.

Baker, D. N., et al. 2013. A long-lived relativistic electron storage ring embedded in Earth's outer Van Allen belt. *Science* 340:186–190.

Li, W., et al. 2014. Radiation belt electron acceleration by chorus waves during the 17 March 2013 storm. *Journal of Geophysical Research Space Physics* 119:4681–4693.

Morley, K., M. G. Henderson, G. D. Reeves, R. H. W. Friedel, and D. N. Baker. 2013. Phase space density matching of relativistic electrons using the Van Allen Probes: REPT results. *Geophysical Research Letters* 40:4798–4802.

Mozer, F. S., O. Agapitov, V. Krasnoselskikh, S. Lejosne, G.D. Reeves, and I. Roth. 2014. Direct observation of radiation-belt electron acceleration from electron-volt energies to megavolts by nonlinear whistlers. *Physical Review Letters* 113:035001.

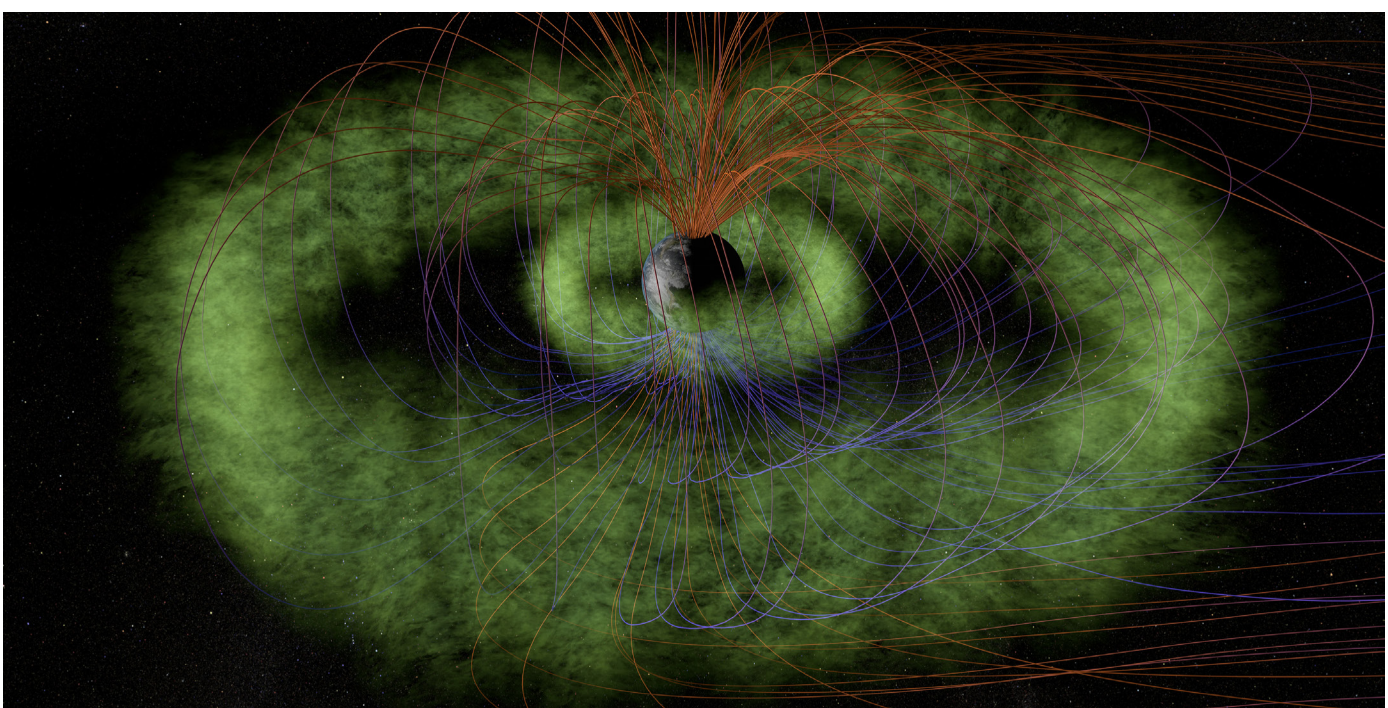
Spence, H. E., et al. 2013. Science goals and overview of the energetic particle, composition, and thermal plasma (ECT) suite on NASA's Radiation Belt Storm Probes (RBSP) mission. *Space Science Reviews* 179:311–336.

Turner, D. L., et al. 2014. Competing source and loss mechanisms due to wave-particle interactions in Earth's outer radiation belt during the 30 September to 3 October 2012 geomagnetic storm. *Journal of Geophysical Research Space Physics* 119:1960–1979.

Turner, D. L., et al. 2014. On the cause and extent of outer radiation belt losses during the 30 September 2012 dropout event. *Journal of Geophysical Research Space Physics* 119:1530–1540.

For relevant Web links, consult this issue of *American Scientist Online:*

http://www.americanscientist.org/issues/id.110/past.aspx



NASA

Fifty Years of Earth-observation Satellites

Views from space have led to countless advances on the ground in both scientific knowledge and daily life

Andrew J. Tatem, Scott J. Goetz and Simon I. Hay

The former Soviet Union's launch of the world's first artificial satellite, Sputnik 1, heralded the era of satellite remote sensing. Since that epic moment on October 4, 1957, hundreds of Earth-observing satellites have followed. Half a century of imagery has provided both iconic views and unprecedented scientific insights.

The science of satellite remote sensing integrates the understanding, interpretation and establishment of relations between natural phenomena and measurements of electromagnetic energy that is either emitted or reflected from the Earth's surface or its atmosphere. These measurements are made for a large number of locations on the Earth's surface by sensors onboard spaceborne satellites and are output as imagery. The 50 years since the first satellite was launched have seen spaceborne remote sensing advance from the small-scale production of low-resolution images for a select few, motivated primarily by military requirements in the Cold War era, to the daily acquisition of over 10 terabytes of information, increasingly available to all, motivated largely by the needs of Earth-observation science.

More than 150 Earth-observation satellites are currently in orbit, carrying sensors that measure different sections of the visible, infrared and microwave regions of the electromagnetic spectrum. The majority of Earth-observation satellites carry "passive" sensors, measuring either reflected solar radiation or emitted thermal energy from the Earth's surface or atmosphere. Newer satellites also employ "active" sensors that emit energy and record the reflected or backscattered response, from which information about the Earth can be inferred.

The features of the instruments depend on the purpose for which each was designed, varying in several aspects. In simple terms, these are: the minimum size of objects distinguishable on the Earth's surface (spatial resolution), the size of the region of the electromagnetic spectrum sensed (spectral extent), the number of digital levels used to express the data collected (radiometric resolution) and the intervals between imagery acquisition (temporal resolution). Moreover, the number of regions of the spectrum for which data are collected, the time taken to revisit the same area of the Earth, the spatial extent of images produced, and whether the satellite's orbit follows the Sun-illuminated section of the Earth (Sun synchronous) or remains over a fixed point on the Earth (geostationary) all vary between satellites and their sensors. The development of satellites over the past 50 years has also been in step with increasing computing capabilities. As data storage capacities and processing speeds increase, so has the ability of Earth-observation satellites to capture, process and return information.

Taking Off

Although the first images of Earth from space were actually taken in 1946 over the New Mexico desert from a camera attached to a V-2 rocket, the era of satellite remote sensing began with Sputnik 1, which completed an orbit of the Earth every 96 minutes and transmitted radio signals that could be received on Earth. This success was followed by Sputnik 2 a month later, in November 1957, and by the first U.S. satellites, Explorer 1 in January 1958 and Vanguard 1 in March 1958. Vanguard 1 remains the oldest satellite still orbiting the Earth and produced the first upper-atmospheric density measurements. The first satellite designed specifically for Earth observation was Vanguard 2, but technical problems meant that it collected little of the intended data on cloud cover. It was superseded by TIROS-1 in 1960, which produced the first television footage of weather patterns from space.

The success of TIROS-1 led to a stream of meteorological satellites and also provided the basis for subsequent development of devices designed specifically for land observation. The National Oceanic and Atmospheric Administration (NOAA) series of satellites followed the TIROS satellites and carried an instrument called the Advanced Very High Resolution Radiometer (AVHRR), which measured the reflectance from Earth in five spectral bands, ranging from the visible to the infrared. Although it was designed for meteorological purposes, this sensor proved successful for land and sea

Andrew J. Tatem holds a joint appointment as a researcher at the Spatial Ecology and Epidemiology Group of the Department of Zoology at the University of Oxford, and as a research associate at the Centre for Geographic Medicine in Nairobi, Kenya. He received his Ph.D. in 2001 jointly from the Departments of Electronics and Computer Science and Geography at the University of Southampton. Scott J. Goetz is a senior scientist at the Woods Hole Research Center and an adjunct associate professor at the University of Maryland, College Park, where he also earned his doctorate in 1996. Simon I. Hay completed his doctorate in the Department of Zoology at the University of Oxford in 1996 and remains in the same department. He works with the Malaria Atlas Project (www.map.ox.ac.uk), principally funded by the Wellcome Trust, U.K. Address for Tatem: Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK. Internet: andy.tatem@zoo.ox.ac.uk

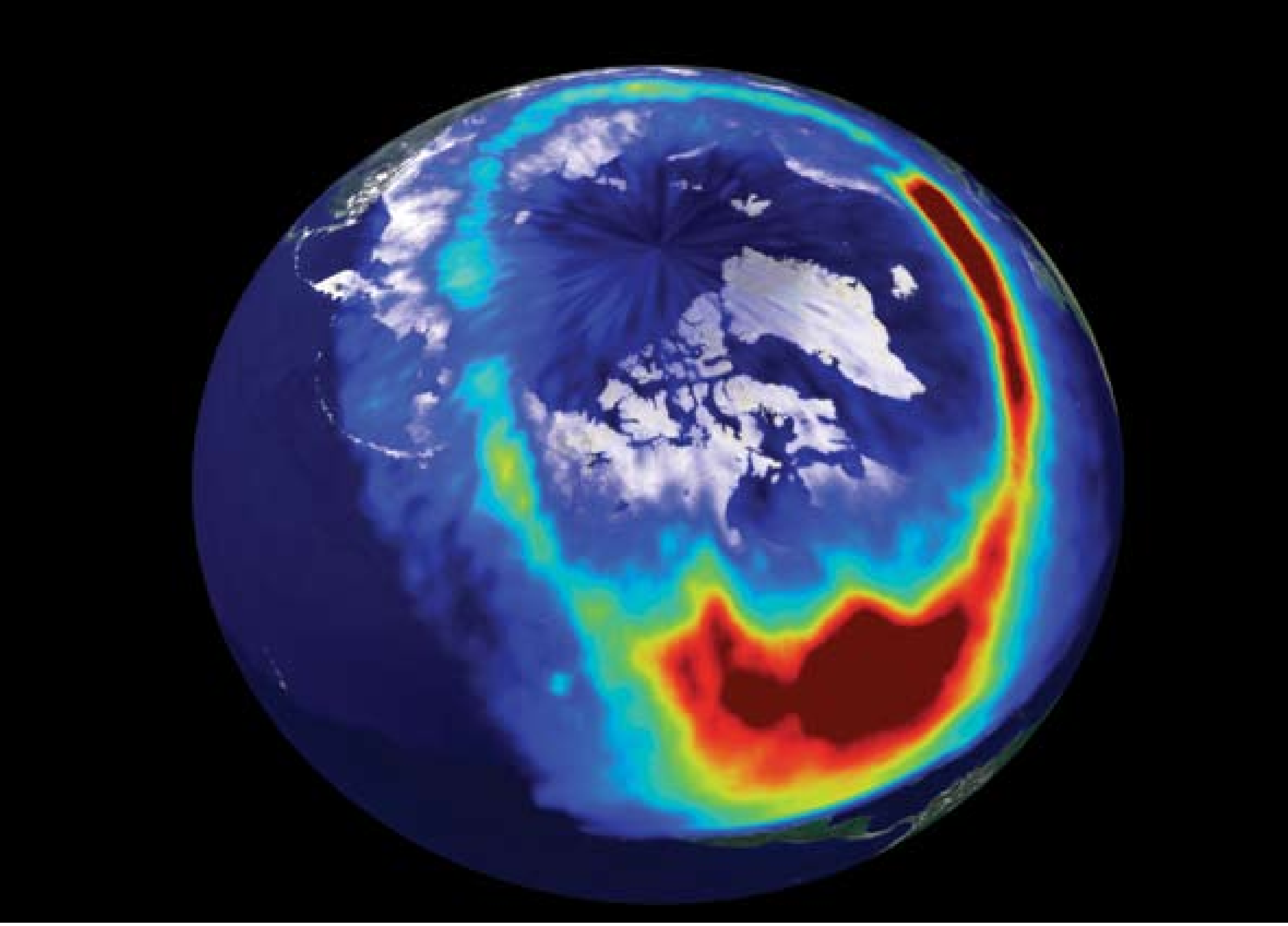


Figure 1. Views such as this one are made possible by satellites orbiting the planet, a feat they have been performing for the past half-century. Although such satellites were initially put in place for military uses, most current ones are used to observe the Earth, and they have provided a wealth of information about the world. This image of an aurora was compiled from data collected in July 2000 by NASA's Polar satellite, which ceased operation earlier in 2008. The data were recorded in ultraviolet light, as the event occurred during daylight hours. False color from blue to red corresponds to increasing magnetic activity. Image is courtesy of NASA/Goddard Space Flight Center and the Scientific Visualization Studio.

observation, providing multitemporal measurements at a global scale.

A key development from 1960 to 1980 was the use of multispectral sensors, stimulated in part by the declassification of military satellites that used both the infrared and microwave bands to observe the Earth's surface. Following pioneering research by the U.S. National Aeronautics and Space Administration (NASA) and the U.S. National Academy of Sciences to assess the utility of Earth observation in forestry and agriculture, NASA launched Landsat 1 in 1972 to monitor Earth's land areas. Landsat images depicted large areas of the Earth's surface in several regions of the electromagnetic spectrum, including both the visible and the near-infrared, and at spatial resolutions useful for many practical applications, such as assessing land cover and use.

Landsat 1 spawned a series of "enhanced" Landsat missions, eventually carrying to orbit the Enhanced Thematic Mapper Plus, capable of capturing data in as many as eight spectral bands, again in the visible to near-infrared, at a spatial resolution of predominantly 30 meters. These missions formed a model for similar land-observation satellites and sensors over the following decades, such as the French *Systeme Pour l'Observation de la Terre* and more recently NASA's Advanced Spaceborne Thermal Emission and Reflection radiometer. The Nimbus satellites, begun in 1964, were also a landmark series, carrying sensors capable of monitoring oceanic biological processes, atmospheric composition and ice-sheet topography. The Nimbus sensors included visible-light cameras, infrared and microwave radiometers, spectrometers, ultraviolet backscatter sensors and coastal-zone color scanners.

The 1980s saw significant advances in the capabilities of existing technologies as well as the development of new ones, including hyperspectral sensors that combine information from several spectral bands, multi-angle spectrometers that combine the views from several azimuths, and spaceborne radar. Active microwave systems have been used for tracking moving objects since the early 20th century, but only in the past two decades have sensors onboard satellites produced active microwave images, where the instruments send out radar pulses and measure their reflectance. Synthetic aperture radar (SAR) is a variation on this technology that can sense through cloud cover and without daylight, measuring the time delay between emission and return, thus es-

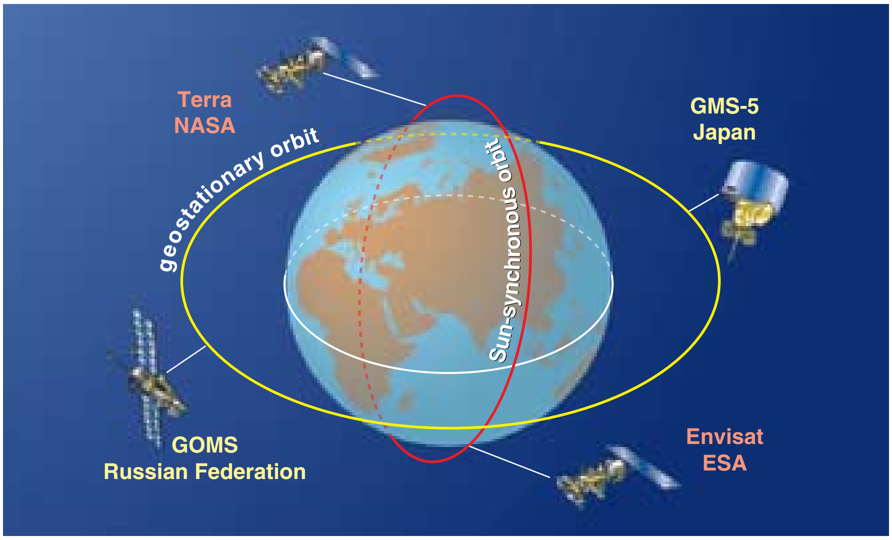


Figure 2. Most of the hundreds of earth-observing satellites occupy one of two types of orbits. A geostationary orbit rotates in sync with the planet, keeping the satellite over a particular location. Examples of satellites in geostationary orbit include Russia's GOMS and Japan's GMS-5, both used for meteorological purposes. A Sun-synchronous orbit passes over the same spot on the Earth at the same time every day; many of these orbits also go over the poles. NASA's Terra satellite and the European Space Agency (ESA) Envisat occupy this orbit. Illustration based on information from the World Meteorological Organization.

tablishing the location, height and scattering properties of the Earth's surface. SAR takes data with its relatively small antenna in multiple positions when it is transmitting and receiving. These signals are combined, accounting for the time delay between them, to give the same information as a much more costly and cumbersome large antenna would be able to provide.

Satellite radar applications have now diversified considerably, with sensor configurations that include altimeters sensitive enough to measure sea-level height with a precision of several millimeters and scatterometers to measure surface roughness. Polarimetric imagers, which detect the relative intensity of the polarized components of reflected radiation, and interferometric imagers, which sense the superposition of different wavelengths, are used to monitor minute land and ice movements. In addition, improved technologies and the continued declassification of military satellites now provide the highest spatial-resolution satellite views of the Earth ever seen, with objects 60 centimeters across distinguishable in images from Quickbird, a privately owned satellite.

The most recently introduced satellite remote-sensing instrument is the laser, principally used for topographic and ice-sheet mapping, but also to measure atmospheric properties and the Earth's surface by fluorescence. Substances such as chlorophyll naturally fluoresce at specific wavelengths, allowing for calculation of the amount of plant life in a certain area, such as in an oceanic algal bloom. Fluorescence is also useful in studying the atmosphere. NASA's Calipso satellite uses green and infrared lidar, or laser pulses, to measure the backscattered reflectance, or fluorescence, of clouds, which gives information not only about the altitudes of clouds but also the properties of aerosols within them. For instance,

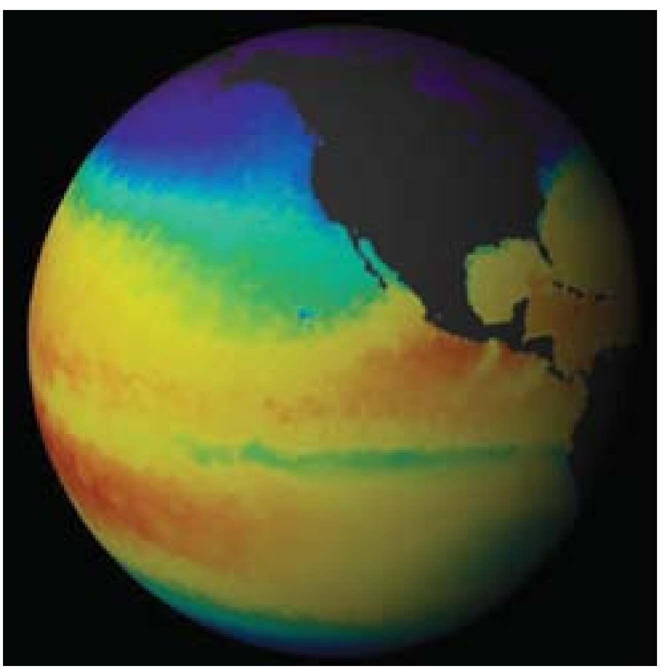


Figure 3. A 1999 image of the El Niño Southern Oscillation (ENSO) event, from the radiometer aboard ESA's ERS-2 satellite, shows rising sea-surface temperature in the Pacific Ocean. This image captures the period when ENSO currents were switching to colder "La Niña" conditions. Image courtesy of ESA.

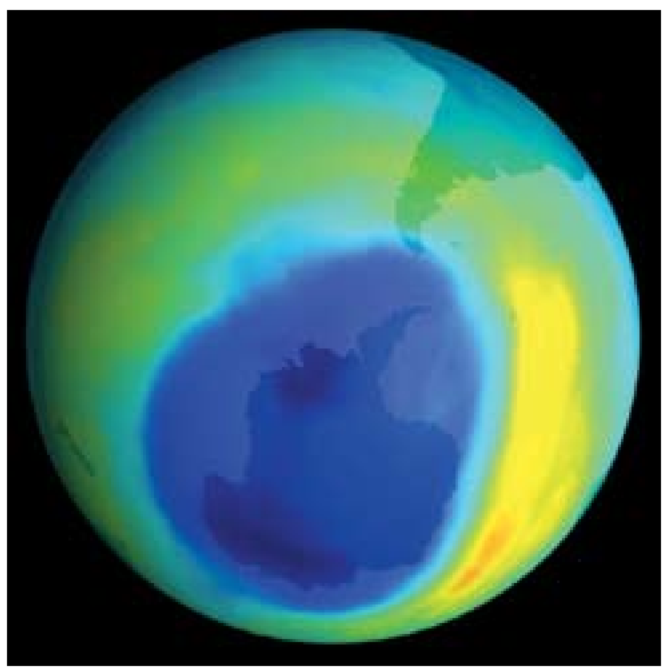


Figure 4. Although it's not really a "hole," there is an "ozone depleted area" over the Earth's southern pole. Taken in September 2000 by NASA's Earth Probe satellite, this image shows the area at 11 million square miles. Image courtesy of the TOMS science team and the Scientific Visualization Studio, NASA/GSFC.

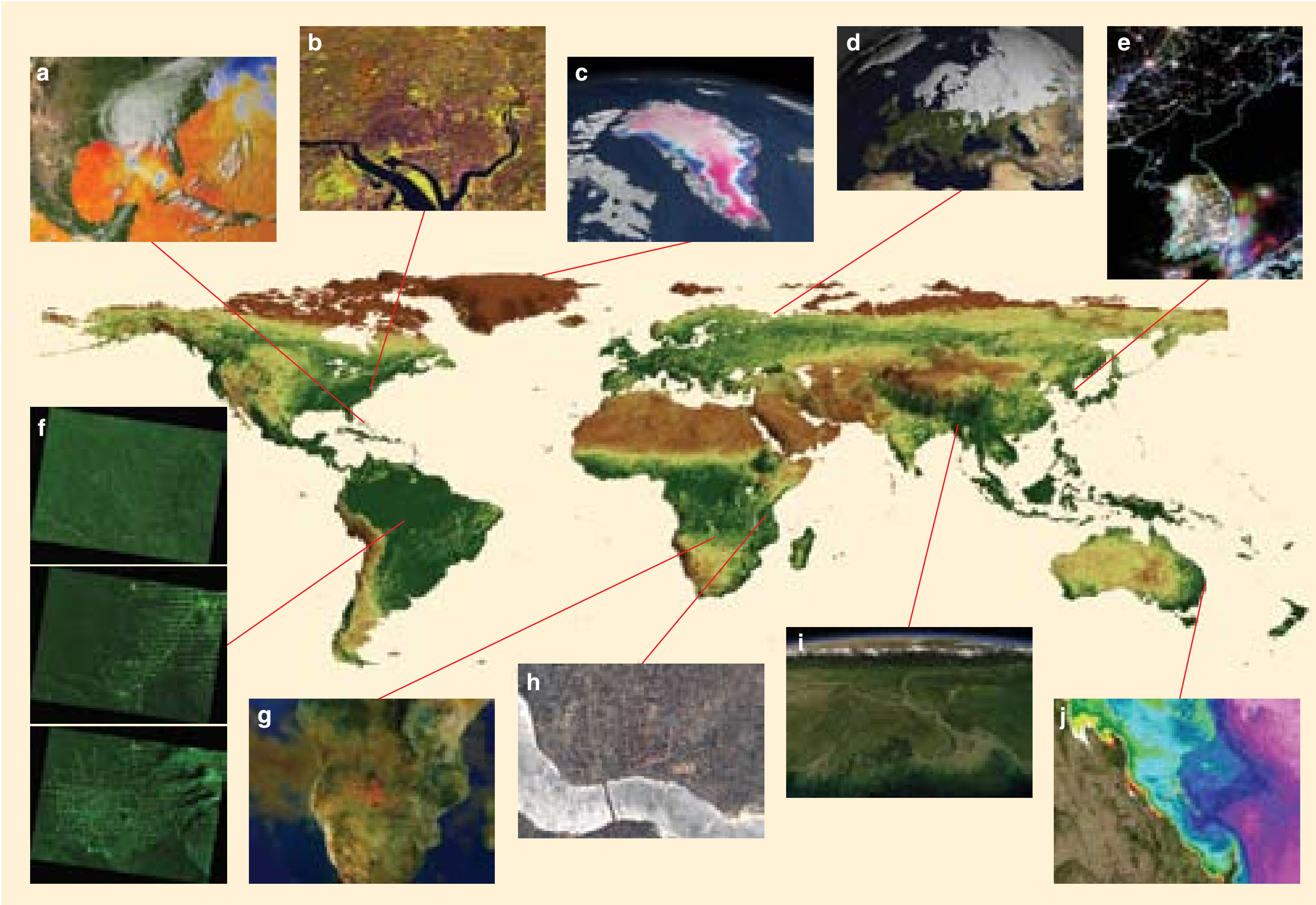


Figure 5. Satellites provide a wealth of information across the globe and in multiple measurement bands. A compilation of satellite images taken in multiple spectral bands gives a whole-Earth picture of land cover, called the normalized difference vegetation index (NDVI) *(center)*. Other images show Hurricane Katrina and sea-surface temperatures as seen by the Terra satellite's Moderate Resolution Imaging Spectrometer (MODIS) *(a)*; a view of Washington, D.C., in false color from Landsat *(b)*; recent Greenland ice sheet elevation changes, red showing an increase and blue a decrease, from SeaSat data *(c)*; European snow coverage from MODIS imagery *(d)*; the stark difference in nighttime lights between North and South Korea, as shown by Defense Meteorological Satellite Programme Operational Linescan System data *(e)*; the progression of Brazilian rainforest loss from Landsat images taken in 1975, 1992 and 2001 *(f, from top to bottom)*; wildfires and smoke in South Africa, as measured by the TIROS and Nimbus satellites *(g)*; the Kenyan coastal resort town of Kilifi, from IKONOS, a commercially owned satellite *(h)*; a Landsat-ETM view of Bangladesh and the Himalayas, with vertical exaggeration applied *(i)* and chlorophyll concentrations off the northeast coast of Australia, from MODIS data *(j)*. Images are courtesy of the authors and the Scientific Visualization Studio, NASA/GSFC.

Calipso spotted a large sulfur dioxide plume that would not have been visible to many other sensors.

Since the early 1990s, two diverging trends in satellite design and operation have developed. First, the large national space organizations, including both NASA and the European Space Agency (ESA), have focused their Earth-observation resources principally on the design and launch of large multisensor platforms, with each sensor designed to monitor a specific aspect of Earth-system processes, frequently at the global scale. Launched in December 1999 and May 2002 respectively, Terra and Aqua are the first of a series of multi-instrument spacecraft forming NASA's Earth Observing System. The next one in the works is the National Polar Orbiting Environmental Satellite System (NPOESS) Preparatory Project, designed as a "bridge mission" to provide a link between the current Terra and Aqua platforms and the next-generation NPOESS mission, currently scheduled for launch in 2013. Additionally, March 2002 saw the launch of ESA's Environmental Satellite, which carries 10 different sensors; at the size of a double-decker bus, it is the largest Earth-observation satellite ever built.

A second, contrasting trend in satellite design is toward smaller, cheaper national satellites. More than 20 countries are now either developing or operating remote-sensing satellites. Typically these are modeled after the Landsat design. Since instrument and launch costs have fallen, lower-income countries such as India, Brazil and Nigeria have launched their own Earth-observation satellites. Many of these new satellites are developed and launched by commercial operators, and are capable of collecting images on demand for a per-item fee, in a variety of operational modes.

Today's Busy Space

Fifty years of Earth-observation satellite development has provided a wealth of memorable images and has driven forward our understanding of Earth-system processes. Today satellite observations are significant data sources for monitoring, measuring and understanding the Earth's terrestrial, aquatic and climatic environments, as well as how they are changing and how each reacts to human influence. Some of the most revolutionary advances brought about by 50 years of remote-sensing progress have been in improving and updating maps.

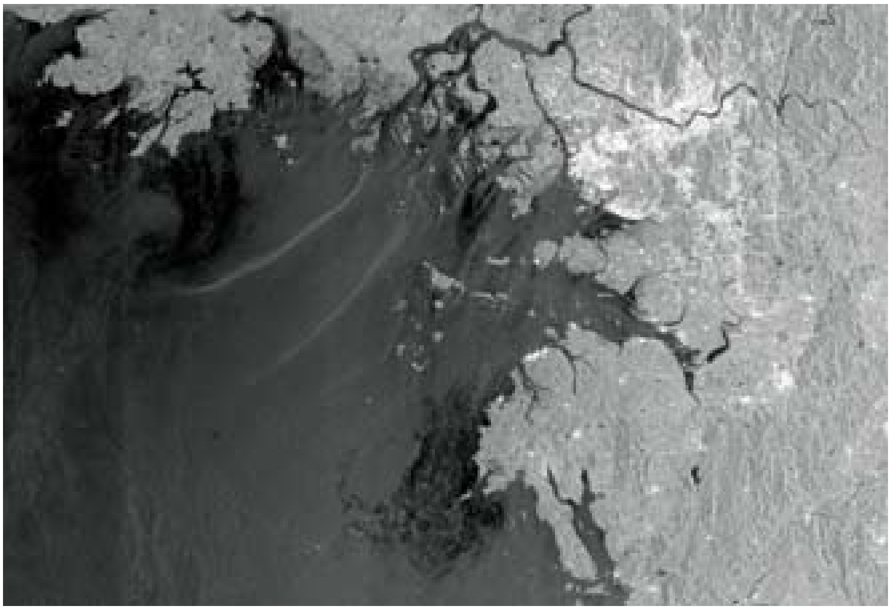
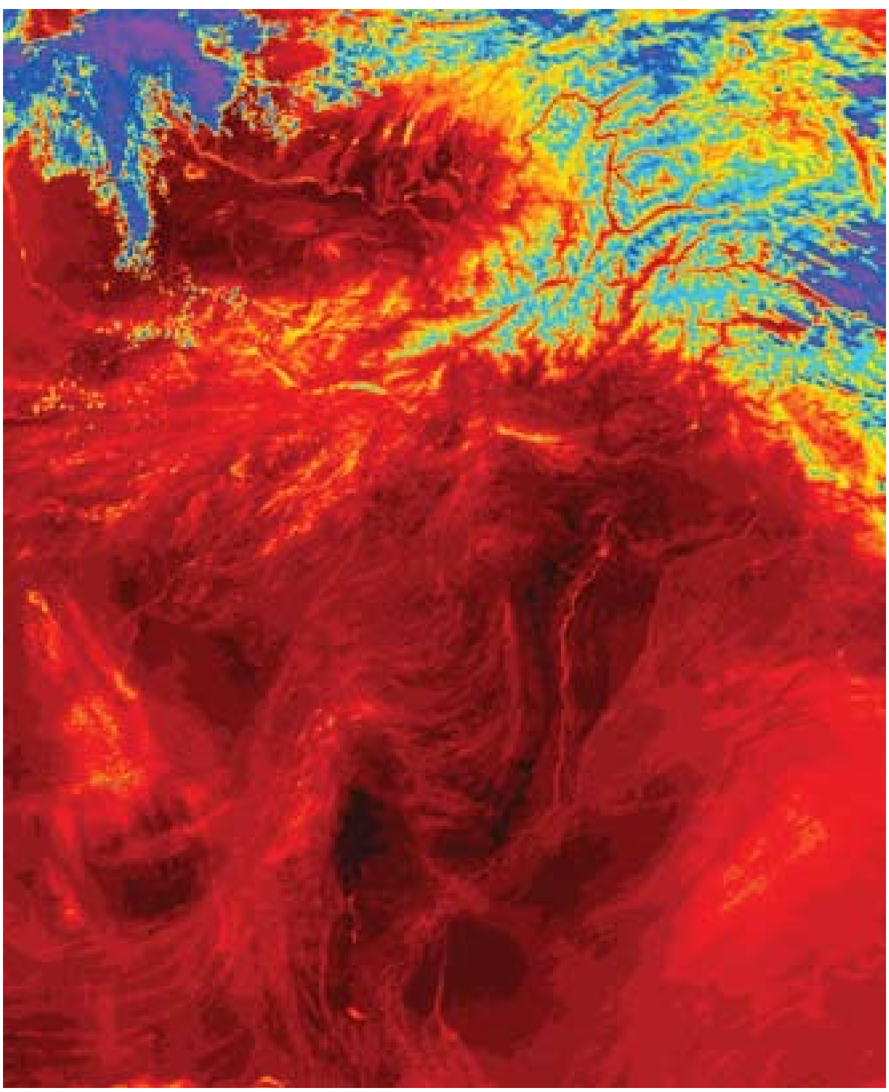


Figure 6. In December 2007 the Hong Kong-based *Heibei Spirit* tanker, at anchor off the coast of South Korea, was punctured by a crane-carrying barge that broke free of its towing tugboat, creating a huge slick of more than 2.5 million gallons of crude oil. The spill *(dark area)* was captured by the synthetic aperture radar instrument aboard ESA's Envisat. Image courtesy of ESA.



From the first basic satellite-derived land-cover maps of the 1960s, to today's stunning online three-dimensional replicas of the Earth, cartography based on satellite imagery has proved to be a consistent and repeatable approach. Such imagery has changed the paradigm of mapping, moving it beyond political borders and topographic landscapes. By sensing outside the visible spectrum, satellites have given us the first large-scale maps of weather patterns, vegetation health, atmospheric pollutants, soil moisture and rock types, among others. Moreover, satellite-derived cartography of the Earth's climate regions and habitats has helped to map species distributions (from tsetse flies to elephants) and disease risks (from Ebola to malaria).

Since the 1940s, the interpretive use of aerial photography for geological and land-cover mapping and evaluation has been widespread, providing an efficient and low-cost approach for resource allocation and for targeting key areas for ground-based surveys. The advent of satellite imagery added further advantages by introducing digital processing, allowing larger areas to be viewed in single scenes, and enabling the combination of visible-light images with a variety of compatible imagery types, such as topography and radar. Satellite imagery is also much easier to update and refine, although it has yet to reach the sub-centimeter spatial resolution of aerial photographs.

Not surprisingly, then, some of the earliest scientific advances based on satellite observations came in the field of geology, where mineral and energy exploration, waste disposal and tectonic modeling all took advantage of the new data sources. For instance, in waste disposal, satellite imagery has been used to locate ideal sites, to detect contaminated land or illegal waste burial and to identify potential fault lines that could allow seepage of waste into groundwater.

Additionally, multispectral measurements significantly improved land-cover assessments as the reflectance from different regions of the spectrum could be combined into indices, such as the

Figure 7. NASA's Aqua satellite recorded this image of the area around the Indus River during a heat wave in May 2004. Land temperatures peaked at 153 degrees Fahrenheit. Blue at the top of the image shows frozen peaks in the Himalaya mountains, in sharp contrast to the deep red of the scorching valleys below. Image is courtesy of Jacques Descloitres, MODIS Rapid Response Team, NASA/GSFC.

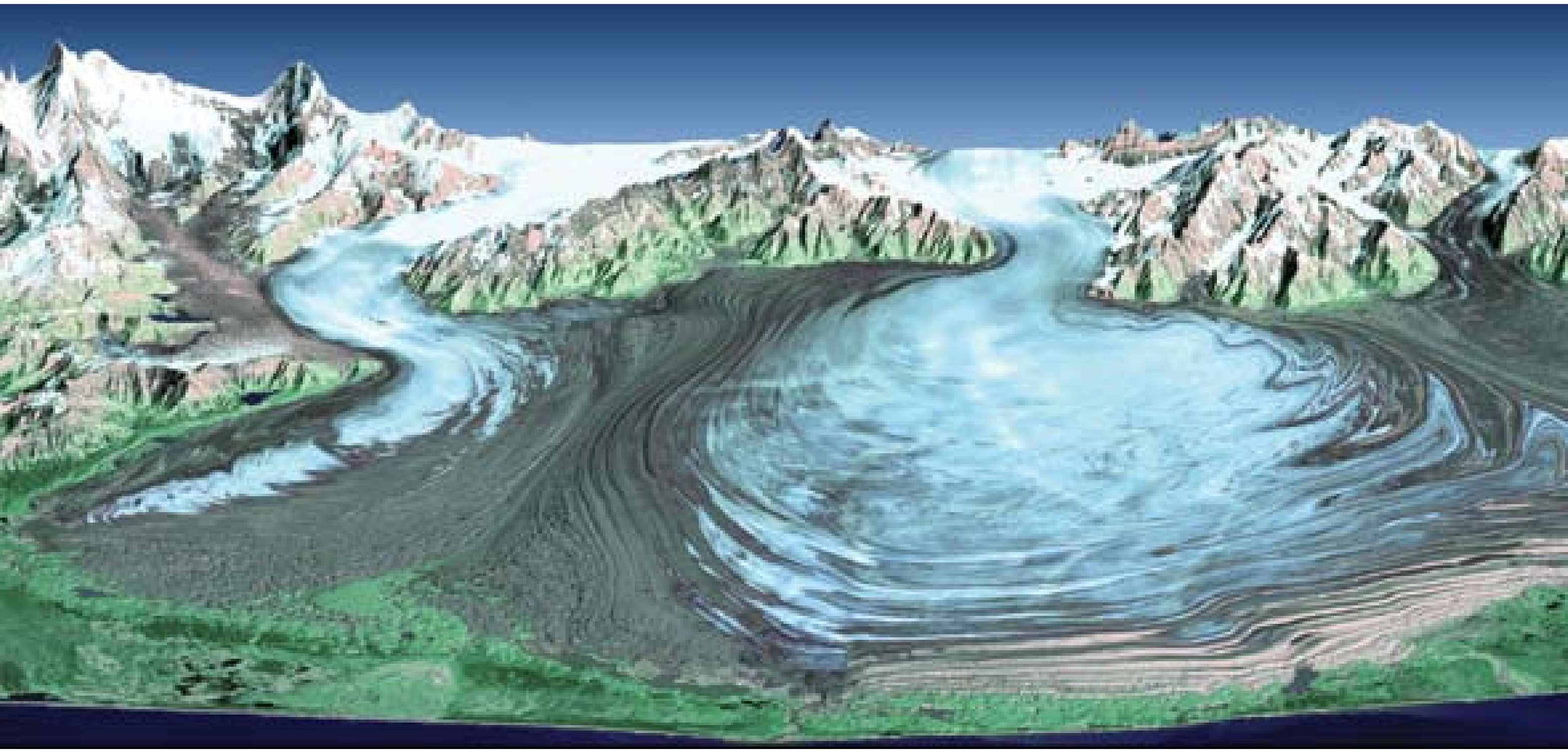


Figure 8. Malaspina Glacier in southeastern Alaska is a classic example of a piedmont glacier, where valley glaciers exit a mountain range onto a broad lowland and spread out. A unique perspective was created by combining a Landsat image, made with both visible and infrared light, with an elevation model from the Shuttle Radar Topography Mission aboard the Space Shuttle Endeavor in 2000. Image courtesy of NASA/JPL/NIMA.

normalized difference vegetation index (NDVI), which exploits the fact that healthy vegetation absorbs light in the red part of the spectrum but strongly reflects near-infrared radiation. The unique multispectral reflectance signatures of each type of surface on the Earth could also be quantified and exploited for accurate and automated mapping.

Although efficient land-cover classification approaches were developed and refined for mapping based on Landsat imagery, it was the AVHRR and its more frequently acquired imagery that provided unprecedented insights about our changing planet. Weekly imagery from the sensor provided the first views of the dynamics of land cover, biomass and primary production across entire continents. Analysis of the long time series of AVHRR imagery, along with an improved understanding of the relations between electromagnetic-energy reflectance and ecological features, made possible the study of ecology on a global scale. These findings, among many others, gave the first quantification of the impacts of the El Niño Southern Oscillation (ENSO) on African crop and livestock production. In addition, the data have shown some unexpected trends in the so-called "greening of the north" phenomenon, where plant productivity in northern high latitudes was thought to be on the rise due to a longer growing season. Greening does continue in tundra regions, but it turns out actually to be on the decline in boreal forest because of hotter, drier air masses over continental interiors.

As archives of Landsat imagery have built up over the years, so have more-detailed insights into land-cover changes, exemplified by large-scale mapping of deforestation, useful not only for land-use planning but also for screening for such activities as illegal logging. The

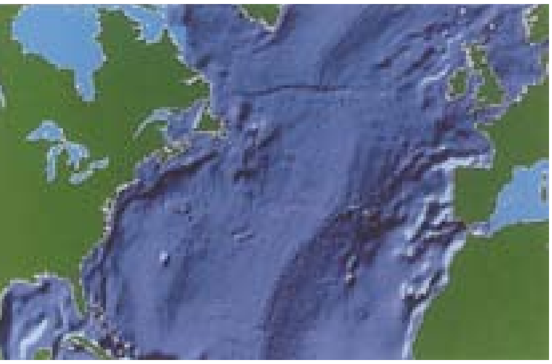


Figure 9. The surface of the ocean is not flat, but contains hills and valleys that echo the shape of the ocean floor over which it flows. ESA's ERS-1 radar altimeter recording over the North Atlantic shows the mid-ocean ridge and continental shelves. Image courtesy of Carel Wakkers, TU Delft, the Netherlands, and ESA.

Figure 10. Storms in the Sahara desert often blow copious amounts of sand and dust out to sea. The Cape Verde islands, about 300 miles off the western coast of Africa, can experience violent dust storms from this distant source, as captured by NASA's Terra satellite in 2000. Image courtesy of Liam Gumley, MODIS Atmosphere Science Team.

Figure 11. Before-and-after images provide an essential resource for understanding the extent of disasters. Images such as this one of New Orleans before Hurricane Katrina *(left)* make it clear just how extensively the city was flooded in a second image 17 days after the storm *(right)*. NASA's Terra satellite captured both of these images, covering an area 10 kilometers by 7 kilometers. Image courtesy of NASA/JPL.

advantages of satellite remote sensing for mapping had similar impacts on soil, agricultural and forestry sciences. Some examples include continental-scale mapping of fires and the advent of precision agriculture and forest management, where growth, water stress, disease and pests can be monitored.

Oceanographic research has also been revolutionized by satellite-based measurements. Researchers can now rapidly acquire and analyze global data sets on sea-surface temperature, surface wind speed and direction, height of surface swells, concentrations of phytoplankton and suspended sediments, wave distributions, and changes in sea-surface height associated with tides and currents. Prior to the 1980s such properties could only be determined through expensive and extensive marine expeditions, but the regular availability of such measurements from spaceborne sensors has now led to long-term studies of sea-level rise and surface-temperature variations, such as the ENSO. Some of the earliest significant advances came from Nimbus-7's Coastal Zone Color Scanner and its pioneering large-scale data collection of oceanic biological processes. Later, the Sea-viewing Wide Field-of-View Sensor (SeaWiFS) provided unprecedented measurements of the response of oceanic biological processes to ENSO and agricultural runoff.

Oceanic phytoplankton contributes around half of the biosphere's net primary production of biomass and therefore represents a significant component of the global carbon cycle. Measurements of chlorophyll distribution from satellites provided the basis for the first large-scale estimates of oceanic net primary production and the discovery of its close coupling to climate. The development of satellite altimeters also enabled global mapping and a new understanding of a range of features through the detection of changes in water height that indicate gravitational concentrations. These include sea-floor topography, tidal-energy dissipation and sea-level rise, as well as detailed characterization of the December 2004 Sumatra tsunami.

Looking from Above

More than 100 satellites have been launched solely for monitoring the Earth's atmosphere. Half have been designed to support weather forecasting, whereas the others have been more research focused. Short-term weather-prediction science has advanced significantly through the use of active microwave instruments, as these operate through cloud cover and without daylight. Microwave and infrared sensors can now be used to map atmospheric temperature profiles, water vapor distribution, surface pressure and precipitation. The Tropical Rainfall Measuring Mission (TRMM) satellite launched in 1997 carries various microwave instruments for precipitation monitoring. TRMM data have contributed to an increased understanding of tropical rainfall processes, including quantification of the inhibiting effects of air pollution on rainfall. As with many satellites initially launched for research purposes, the success of TRMM has meant that its mission has been extended annually well past its expected life.

The interactions of electromagnetic waves with the Earth's atmosphere are determined by both their wavelength and by the atmosphere's pressure and temperature and the particulates suspended within it. The scattering, emission, refraction and absorption of electromagnetic waves interacting with the atmosphere is a complex science, but Earth-observation satellite data have formed the basis of some significant advances in this realm, including the first global measurements and maps of the Arctic and Antarctic ozone "holes," through use of the Nimbus-7 Total Ozone Mapping Spectrometer to measure backscattered solar ultraviolet radiation. The same sensor was used to quantify global tropospheric ozone levels related to air pollution, whereas improved sensors provided unprecedented maps of global smoke, dust and nitrogen oxide levels.

The study of the mechanisms controlling the global climate system and its changes has become heavily dependent on the use of satellite observations. The data are routinely used to populate models of climate, but they also both confirm model results and provide new data that either contradict predictions or indicate where models fall short.

Satellite remote sensing has proved invaluable in studying the Arctic and Antarctic without the need for humans to disturb or endure these fragile, extreme environments. Remote sensing of the cryosphere is, however, sometimes restricted by the polar environment. The orbital inclination of many satellites means that their sensors do not cover regions with latitudes greater than 80 degrees. Moreover, at any time, at least 50 percent of the polar regions are covered

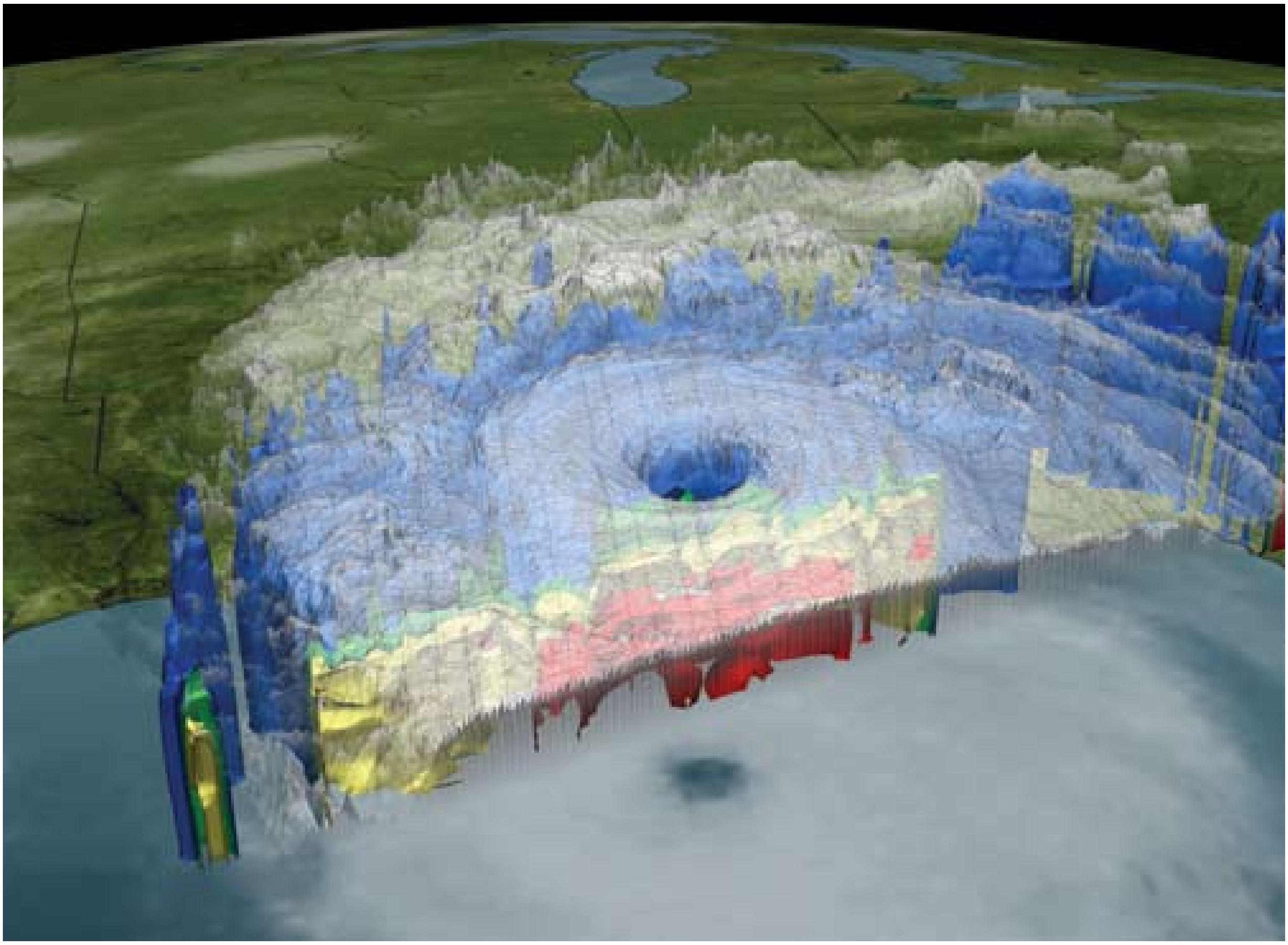


Figure 12. A composite view of Hurricane Katrina on August 28, 2005, shows data from several instruments aboard two satellites. The Tropical Rainfall Measuring Mission satellite, a joint venture of NASA and the Japan Aerospace Exploration Agency, looks underneath a storm's clouds to reveal the underlying rain structure. Blue indicates at least a quarter inch of rain an hour, whereas green represents half an inch, yellow an inch and red two inches. The Geostationary Operational Environmental Satellite, run by NASA and the National Oceanic and Atmospheric Administration, provided visible meteorological data. Image courtesy of NASA/JAXA.

by cloud, and during their respective winters each endures extended periods of darkness, making the consistent use of visible and infrared sensors problematic. These issues have led to the extensive use of microwave instruments. The continuous availability of radar data over the past decade has provided significant advances in understanding the cryosphere.

Sea-ice extent and movement are key indicators of climate change, and are also important for ship routing and weather forecasting. A succession of passive microwave radiometers has led to continuous records since 1972, with spatial resolution improving with each new radiometer. At the same time, SAR data have enabled discrimination between seasonal and persistent ice types, and monitoring of sea-ice reductions consistent with global warming. Significant disintegrations of Antarctic Peninsula ice shelves, also coincident with climate warming, were observed using optical and SAR imagery, as was accelerated ice discharge on Greenland.

Ice thickness also represents an important climate-change indicator. Although its measurement is problematic, data from satellite radar altimeters and infrared radiometers have shown promise as model inputs, especially when on-site numbers are available for calibration. Satellite altimeter data have even been used to map a vast freshwater lake beneath Antarctica. Topography remains perhaps the most fundamental observation for an ice sheet, with regular, accurate measurements providing information on direction and magnitude of flows, which are vital parameters for glacial mass-balance estimates. Radar and laser altimeters, as well as SAR interferometry, have all proved capable of producing accurate measurements of ice-sheet topography and dynamics.

Recent years have seen the application of data from Earth-observation satellites extend into new research fields. Urban and regional planners require nearly continuous acquisition of data to formulate policies and programs, and new satellites with increased spatial and spectral resolution provide data to meet these requirements. From flood-risk modeling, subsidence detection and traffic management, to archaeological surveying, landmine detection and even crime-risk mapping from nighttime imagery, satellite imagery is now widely used for societal applications. The 35-year archive of Landsat imagery provides data for land-use and urban-growth modeling, whereas nighttime imagery of electrified urban areas is facilitating the construction of global human-population spatial databases, which are finding applications in disease-burden estimation and epidemic modeling.

Globally consistent satellite data on a range of climatic variables now exist, including temperature, rainfall and vegetation area. These data are beginning to find significant applications across the low-income regions of the world in exploring food security, resource accessibility and the construction of early-warning systems in planning for the effects of crop failure and disease outbreaks. The resultant maps are improving decision making and efficient resource allocation. Moreover, with the climatic and environmental preferences and tolerances of numerous species quantified, the same global imagery is helping to infer present and future distributions for improved conservation planning. From the availability of habitats for giant pandas, to the distributions of malarial mosquitoes, satellite imagery has become an important asset for ecologists and epidemiologists alike.

The Big Picture

The last half-century has seen satellite remote sensing come of age as a multidisciplinary research field, with a balance of theory, practice and operational application. It still faces barriers to becoming a fully global and cross-disciplinary data source, particularly in low-income countries, but in many cases these limitations are being reduced. The continued increase in computing power and decrease in costs are making satellite imagery more manageable and affordable. However, the building of image archives spanning different time periods still requires significant resources.

The increasing number of Earth-observation satellites and the availability of imagery are driving down data costs. Free online databases and open distribution of processed imagery are making many types of data available to all. Although this is a welcome trend, it remains exceptional, with even unprocessed data from numerous satellites not readily available and many operators still charging high fees for imagery.

Software for handling and processing satellite imagery was previously rare, as well as complex and expensive, but is becoming widespread and more user-friendly. Basic software is now, in many cases, cheap or even free, but the most powerful and advanced programs still require costly licenses. Training in the use of satellite imagery has also grown as such data become central to numerous disciplines, but cutting-edge computing, imagery and software often mean that course costs remain prohibitively high for institutions in low-income countries.

Increasingly, limitations in satellite-data applications have shifted from the technology of acquiring the data to the techniques on the ground required to optimally exploit the information within the remotely sensed data. The conventional trade-offs in spectral, spatial and temporal characteristics, which must now be solved by choosing imagery from different satellite sensors, are gradually being made unnecessary by new technology. Forthcoming launches and plans should herald the first images with a spatial resolution under half a meter, high spatial resolution SAR imagery, laser imaging and detailed nighttime data. Improvements in data processing and fusion could help eliminate cloud-obscured and nighttime data loss, and provide multi-image virtual databases for modeling of environmental and social processes. Finally, the declassification of military space technology may well provide valuable new data in the future, just as it has in the past.

There can be no doubt that satellite remote sensing is likely to continue to grow as an operational tool for mapping, monitoring and managing the Earth, as a profit-making entity and as a primary data source for Earth-system science. Existing trends in satellite design are likely to continue, and new ones will emerge, driven by both operational need and profits. Although global issues such as climate change and its effects will continue to provide justification for large multisensor satellites, the design directions in which smaller commercial satellites will head is less clear. The potential for real-time imagery has just begun to be realized, and personalized imagery beamed to handheld devices will soon show users their positions in traffic or current weather at their destinations. To speculate further, the online availability of such imagery could facilitate a real-time or "live" Google Earth. Such a resource potentially enables revolutionary studies involving the global tracking of terrestrial and oceanic life, which could help create, for instance, real-time disease epidemic models, dynamic traffic control and reactive conservation—but it also raises significant security and privacy concerns.

Despite significant proven potential, the future supply of high-quality Earth-observation data for research and other applications remains unclear. For instance, funding cuts in U.S. programs have generated concern over a possible data gap in the Landsat imagery series, and budget overruns have both modified the scope and delayed the launch of the NPOESS project. At a time when unprecedented changes are taking place in the Earth's atmosphere, oceans and land surface, it is difficult to rationalize any scaling back of demonstrably successful and valuable satellite remote-sensing programs. Such examples emphasize the need for multinational cooperation in Earth observation to maintain a consistent supply of global data and ensure another 50 years of continuous measurements, stunning images and a deeper understanding of the Earth.

Bibliography

Behrenfeld, M. J., *et al.* 2006. Climate-driven trends in contemporary ocean productivity. *Nature* 444:752–755.

Goetz, S. 2007. Crisis in Earth observation. *Science* 315:1767.

Goldstein, R. M., H. Engelhardt, B. Kamb and R. M. Frolich. 1993. Satellite radar interferometry for monitoring ice sheet motion: Application to an Antarctic ice stream. *Science* 262:1525–1530.

Kaufman, Y. J., D. Tanre, and O. Boucher. 2002. A satellite view of aerosols in the climate system. *Nature* 419:215–223.

The National Research Council of the National Academies. 2007. *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond.* Washington, D.C.: National Academies Press. http://books.nap.edu/catalog/11820.html

Parkinson, C. L., and D. J. Cavalieri. 2002, A 21-year record of Arctic sea ice extents and their regional, seasonal and monthly variability and trends. *Annals of Glaciology* 34:441–446.

Rairoux, P., *et al.* 2000. Remote sensing of the atmosphere using ultrashort laser pulses. Applied Physics B: Lasers and Optics 71:573-580.

Rogers, D. J., S. E. Randolph, R. W. Snow and S. I. Hay. 2002. Satellite imagery in the study and forecast of malaria. *Nature* 415: 710–715.

Sellers, P. J., B. W. Meeson, F. G. Hall, G. Asrar and R. E. Murphy. 1995. Remote sensing of the land surface for studies of global change: models, algorithms, experiments. *Remote Sensing of Environment* 51:3–26.

Stolarski, R. S., *et al*. 1986. Nimbus-7 Satellite Measurements of the Springtime Antarctic Ozone Decrease. *Nature* 322:808–811.

Tucker, C. J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8:127–150.

Turner, W., *et al.* 2003. Remote sensing for biodiversity science and conservation. *Trends in Ecology and Evolution* 18:306-314.

For relevant Web links, consult this issue of *American Scientist Online:*

http://www.americanscientist.org/issues/id.74/past.aspx

The Two-Faced Moon

Investigators are still struggling to understand why the near and far sides of our celestial neighbor are so fundamentally different

P. Surdas Mohit

Everyone has seen the visible face of the Moon, but have you ever wondered what the other side looks like? In October 1959, the Soviet spacecraft Luna 3 snapped the first ever picture of the lunar far side. Astonishingly, it was revealed as strikingly different from the Moon's more familiar hemisphere.

The side of the Moon that we see contains prominent dark patches. Named *maria* (Latin for "seas") by ancient astronomers, these regions were originally thought to be bodies of water. But people now know that the dark color comes from the kind of rock found there. Luna 3 showed that the far side of the Moon is almost bereft of such features. From this initial observation have proceeded a host of others and a gaggle of models competing to account for them. Almost a half-century later, planetary scientists have made a lot of progress, but a definitive explanation for the Moon's hemispheric asymmetry remains elusive. Here I sketch the outlines of a theory that may help to account for it.

My own contribution to the subject comes from my studies over the past few years of large impact basins on the Moon and elsewhere in the solar system. But as I describe below, this work provides just a small piece of the puzzle. The main findings of relevance to this question have been accumulating for decades, many as spin-offs of America's quest to send astronauts to the Moon.

P. Surdas Mohit earned a Ph.D. in Earth and planetary science from Washington University in St. Louis in 2007 and is now a postdoctoral research associate at the Scripps Institution of Oceanography. His research focuses on viscoelastic deformation in various solar-system bodies. In addition to his scholarly papers, Mohit has published popular articles in, for example, the Christian Science Monitor, *the* Vancouver Sun *and the* Toronto Star. *Address: Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, 9500 Gilman Drive, La Jolla, CA 92093. Internet: pmohit@ucsd.edu*

An Ocean of Magma

The Apollo program, catapulted to prominence by President John F. Kennedy's bold commitment in 1961 to land a man on the Moon "before this decade is out," was one of the most ambitious and expensive scientific endeavors in history. Returning a wealth of samples, surface measurements and satellite observations, these missions dramatically advanced scientific understanding of the Moon and spurred the development of many of the principles that are now applied more generally in planetary science.

Among other things, the Apollo missions enabled investigators to forge the basic theory that explains how the major rock types found on the Moon came to be. The key to this understanding was the realization that a large fraction of the Moon was melted during its formation and that the rocks seen at the surface—and others below whose existence can be inferred—all crystallized from a global "magma ocean," one that may have been 500 or more kilometers deep initially.

The energy required to melt all that rock came from the colossal impact between a Mars-sized body and the proto-Earth. Much of the collisional debris coalesced to form the Moon, which arranged itself like a layer cake, with the denser materials sinking to the bottom and the least-dense constituents—including a prodigious quantity of magma—rising to the top.

As this ocean of magma cooled, minerals crystallized in a sequence dictated by their individual solidification temperatures. The least-dense minerals (such as plagioclase feldspar) floated to the surface, whereas denser ones (such as olivine) sank to the bottom. The rocks that comprise such a compositionally layered body are known to geologists as *cumulates*.

Certain elements, referred to as "incompatible," do not fit easily into the crystal structure of minerals. Iron is mildly incompatible, and heavier elements such as uranium and thorium (the principal radioactive elements on the Moon) are extremely incompatible. Thus, the minerals that crystallized early from the Moon's magma ocean contain low concentrations of iron. Most of the iron, and the other highly incompatible elements, resisted incorporation into minerals until the bitter end. As a result, the liquid remaining during the latest stages of solidification of the magma ocean was highly enriched in iron and radioactive elements.

The majority of the lunar surface—the bright highlands—is composed of a light-colored type of rock called anorthosite, which is mostly made of plagioclase feldspar. The large number of impact craters pockmarking these highlands shows that they are very old. The other major component of the lunar surface is dark basalt. Such rocks solidified from lavas that erupted and then pooled in huge impact basins, forming the lunar *maria*.

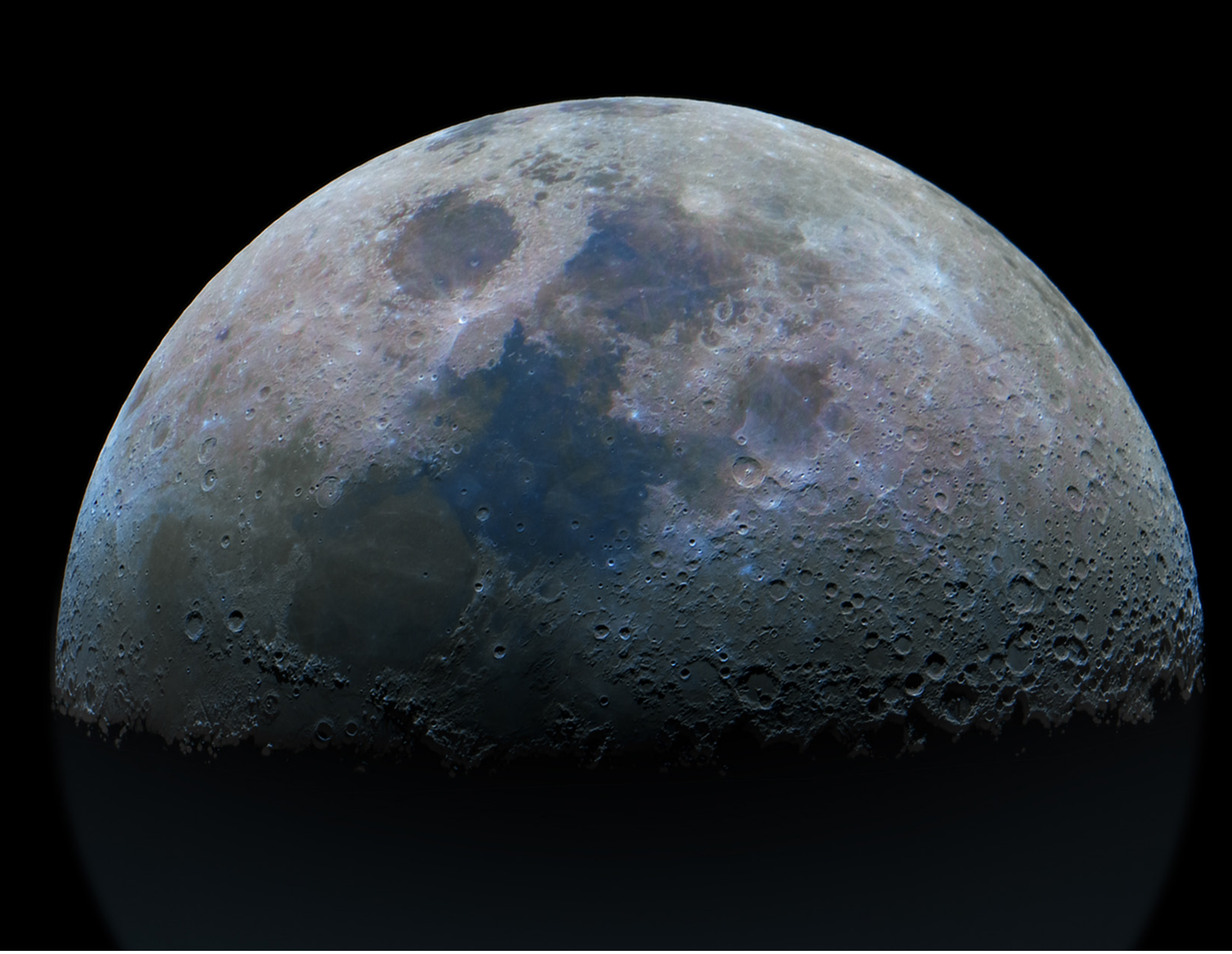


Figure 1. As familiar as its face is, the Moon still presents many scientific mysteries. Some of the most intriguing involve the origin and early history of this nearby celestial body. Key evidence comes from the Apollo missions of the 1960s and '70s, when astronauts visited the Moon's dark *maria* and its light-colored highlands, both of which are easily seen in this telescopic view of the gibbous Moon.

Geochemical analysis of these two rock types reveals the fingerprint of their common origin. The anorthosite contains relatively large amounts of the heavy trace element europium, which is compatible with the crystal structure of plagioclase feldspar. *Mare*-filling rocks show a corresponding depletion in europium, indicating that the source of these basalts was a magma from which the makings of plagioclase feldspar had already been removed.

Equally intriguing was the discovery of a chemical signature known as *KREEP*, which stands for potassium (chemical symbol K), Rare Earth Elements and Phosphorus. Found primarily in the rocks that show signs of being broken up and re-cemented by the shock and heat of impacts, KREEP is extremely rich in incompatible elements. Indeed, KREEP is far more enriched in such elements, including the radioactive elements uranium and thorium, than any terrestrial rock type. This composition is consistent with its having been derived from the very latest stage of the Moon's magma ocean. Most probably, KREEP solidified deep within the Moon, it being found now only at sites where a colossal impact exhumed a portion of the subsurface.

Multiple Asymmetries

The measurements made during the Apollo years and the resulting magma-ocean paradigm helped to explain

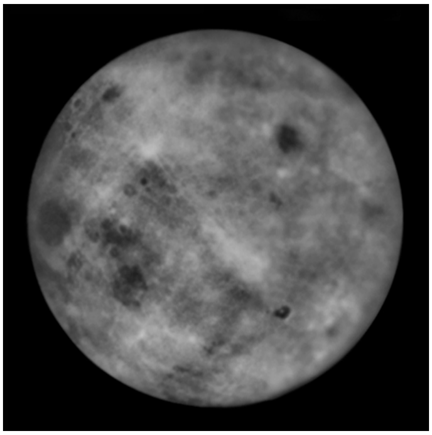


Figure 2. In 1959, the Soviet space probe Luna 3 took the first pictures of the lunar far side. The quality of the images was poor, but they were sufficient to reveal that the formerly invisible hemisphere almost entirely lacks *maria*, which cover a considerable fraction of the near side. The reason for this asymmetry remains the subject of investigation more than four decades later. (Processed mosaic of Luna 3 images courtesy of Ricardo Nunes.)

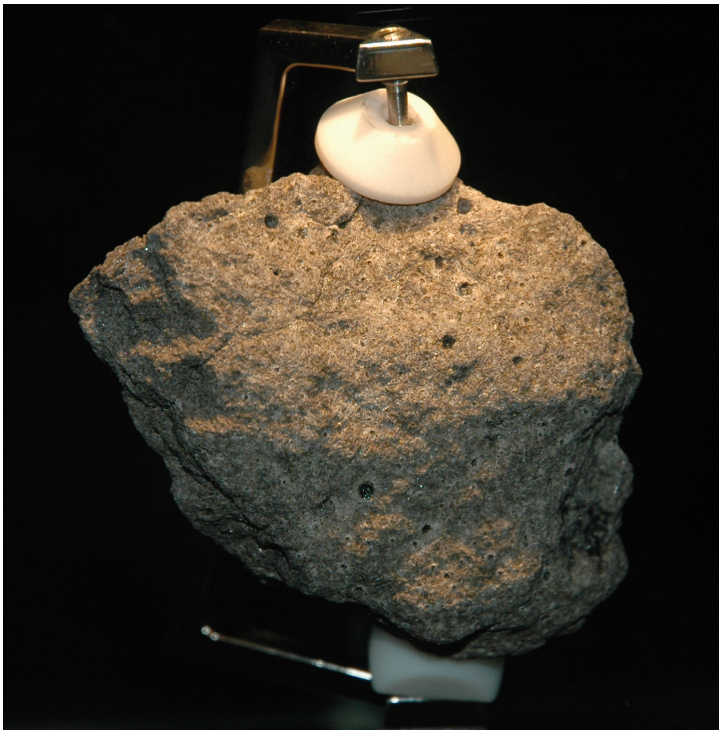


Figure 3. Although many different minerals can be found on the lunar surface, the light and dark coloration of different regions fundamentally reflects just two varieties of underlying rock: basalt *(left)*, which blankets the dark *maria*, and anorthosite *(right)*, which covers most of the remaining lunar surface. (Lunar basalt photograph courtesy of James St. John; lunar anorthosite photograph courtesy of Chip Clark, Smithsonian Institution.)

the patchiness of the Moon, but they were unable to account for why the near and far sides look so different. More insight came in 1994 with the launch of the Clementine spacecraft (also known as the Deep Space Program Science Experiment). A joint project sponsored by the Ballistic Missile Defense Organization and NASA, Clementine provided the first global digital dataset of the Moon. Researchers used the observations Clementine collected, along with those obtained during the 1998 Lunar Prospector mission, to produce global maps of the Moon's topography, gravity and magnetic fields and the abundances of several key elements.

These maps revealed that the near-side–far-side asymmetry has multiple aspects. In particular, the near and far sides were found to differ markedly in the thickness of the crust—the relatively thin, low-density layer overlying the Moon's rocky mantle, which in turn surrounds its metallic core.

It might seem surprising that orbiting space probes could measure the thickness of the lunar crust. In truth, they couldn't. But investigators were able to make some estimates of crustal thickness based on measurements of lunar gravity. The logic they applied

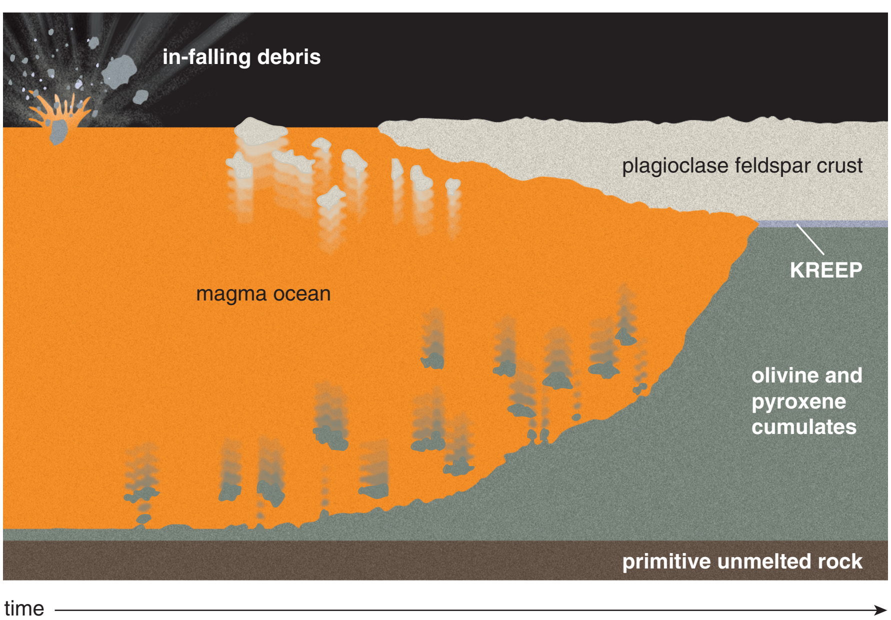


Figure 4. Current understanding of how the light-colored lunar highlands evolved is based on the idea that the accretion of Earth-orbiting debris that formed the Moon generated enough heat to melt much of it, creating a thick ocean of magma, which reached down perhaps 500 kilometers or more. Over time, as the magma ocean cooled, solid minerals formed within it. Minerals that were more dense than their liquid surroundings (primarily olivine and pyroxene) sank to the bottom of the magma ocean, whereas less dense ones, such as plagioclase feldspar, rose to the top, forming the Moon's light-colored, anorthositic crust. Potassium (K) along with Rare Earth Elements (REE) and phosphorus (P), resisted integration into minerals until very late in the process of solidification. As a result, these and similar "incompatible" elements formed into solid rock of a type known as KREEP, which was originally situated in a buried layer between the plagioclase feldspar–rich crust and the olivine- and pyroxene-rich "cumulates" below.

went like this: The variations in topography across the lunar surface have a measurable effect on the force of gravity felt by the spacecraft. The extra mass from the mountainous highlands, for example, produces a slightly stronger downward tug on an orbiting spacecraft, whereas such a probe experiences an unusually weak gravitational pull over lowlands where there is a deficit of mass. It's easy enough to calculate the variation in gravity produced by such changes in topography and subtract it from the measured gravity field. The result then reveals otherwise invisible undulations of the deeply buried interface between the crust and dense mantle beneath.

Specialists combined this information with a few spot estimates of crustal thickness obtained using seismic measurements collected during the Apollo era. In this way, investigators produced a global map of lunar crustal thickness, which proved to be in the range of 35 to 65 kilometers in most places. This map also indicated that the crust on the far side is, on average, substantially thicker than that on the near side.

Clementine and Lunar Prospector data also allowed researchers to chart the concentration of iron at the surface. The resulting maps show much higher iron abundances on the near side. This asymmetry is mostly a result of the large quantity of basalt that erupted and filled the near-side *maria*, although there are also indications that the impacts that formed the near-side basins excavated into a mildly iron-enriched lower crust.

The most prominent of the asymmetries found is for thorium. This radioactive element is concentrated almost entirely in the Oceanus Procellarum region of the central near side.

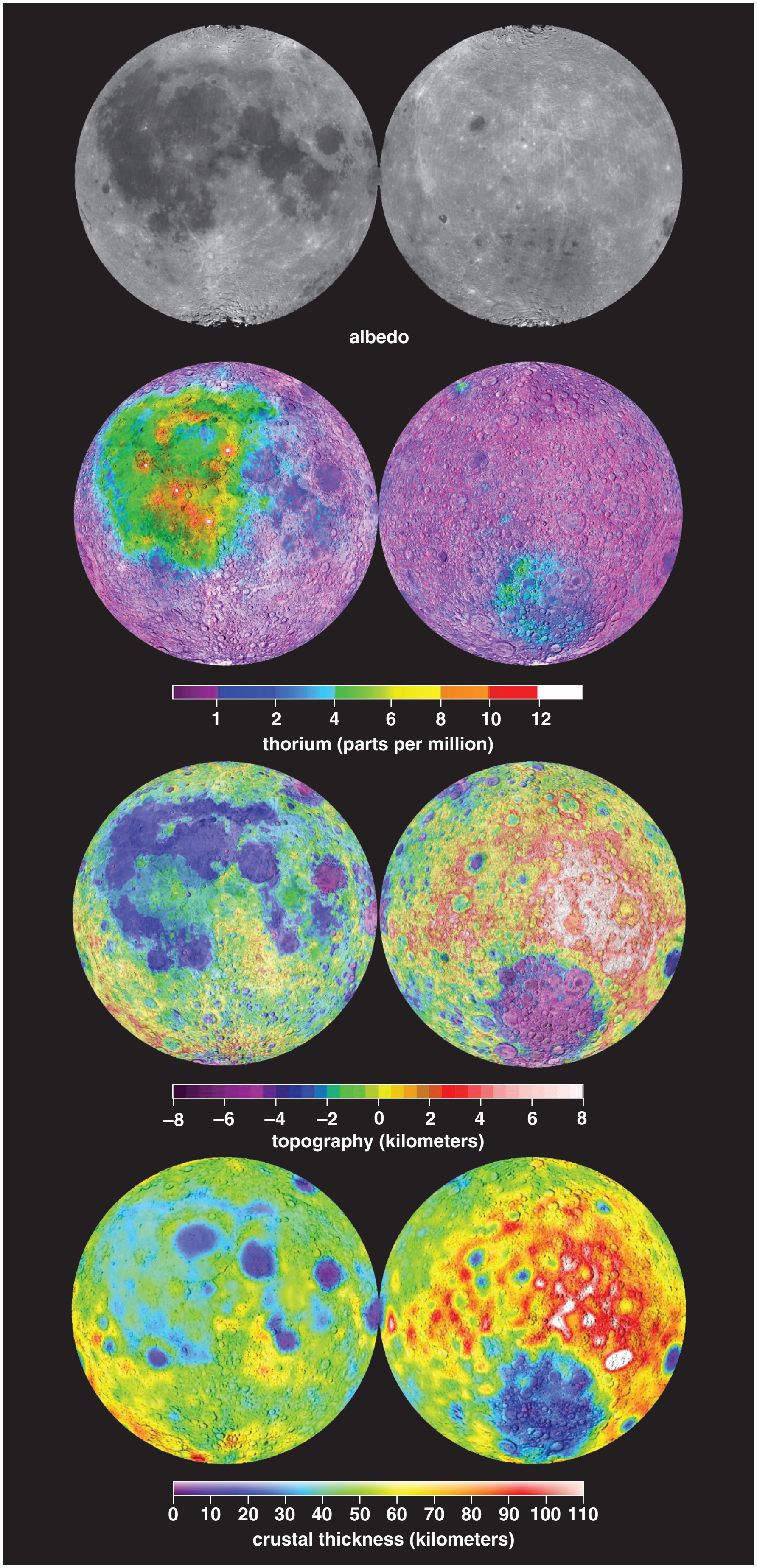


Figure 5. Many properties of the lunar far side contrast markedly with those of the near side. The most obvious is the lack of dark *maria*, a difference that can be seen clearly in the albedo (reflectance) measurements the Clementine spacecraft collected in 1994 *(top)*. The distribution of thorium is very different, too, with most of this and other radioactive elements concentrated in surface rocks of the near side *(second from top)*. Topography and crustal thickness *(second from bottom* and *bottom)* also differ significantly between the near and far sides. (Crustal thickness map courtesy of Mark A. Wieczorek; other maps from www.spudislunarresources.com courtesy of Paul Spudis and the Lunar and Planetary Institute.)

Figure 6. One way to probe the lunar interior is by studying the evolution of large impact basins. Whereas small impacts merely excavate the surface, creating simple, bowl-shaped craters *(left)*, larger ones cause the shocked target rock to lose strength and act temporarily like a liquid. This phenomenon often leads to the appearance of central peaks *(middle)* and, for the largest impacts, creates "multi-ring" structures *(right)*.

A Sea Change

Researchers with an interest in the Moon have advanced several ideas for how these asymmetries came about. To understand these explanations, it's important to look a little more closely at how the *maria* originated.

The Apollo astronauts brought back many samples of *mare* basalt, and subsequent analyses revealed important clues to their formation. These rocks are very different from terrestrial basalts, in that they are both very dense and very rich in iron (and sometimes titanium). This composition suggests that they formed at a late stage during the solidification of the magma ocean.

Geochemists have pieced together the following picture of how the *maria* came to be filled with basalt. After about 95 percent of the magma ocean crystallized, the rocks that subsequently formed from it were very high in iron and contained a large proportion of the mineral ilmenite, a titanium oxide. These ilmenite-rich cumulates were very dense, far more so than the underlying mantle. As a result, they tended to sink over time, as the mantle deformed plastically beneath them.

Because they solidified so late from the magma ocean, the ilmenite-rich rocks incorporated a large quantity of incompatible elements beside iron and titanium. In particular, they took up large amounts of radioactive uranium and thorium, along with potassium, one common isotope of which is also radioactive. As a result, the energy produced by radioactive decay—a significant source of heat over geologic timescales—warmed these rocks, which as a consequence expanded and became less dense. So after first sinking, they eventually became buoyant, rising to the surface and melting. The resultant lavas filled in many of the impact basins, forming the dark *maria*.

A group of scientists at the Massachusetts Institute of Technology and Brown University, led by Shijie Zhong (now at the University of Colorado), have modeled these movements numerically. These investigators showed that, under the right conditions, this process could produce a flow pattern in which the upwelling of hot cumulates occurred only beneath one hemisphere of the Moon.

David R. Stegman and his colleagues at the University of California, Berkeley, later demonstrated that if such upwelling of mantle materials extended down all the way to the lunar core, it might stimulate convective cooling of the liquid-metal core, which (if sufficiently strong) could in turn generate a magnetic field. There is evidence that the Moon once had an internally generated magnetic field, although the matter remains controversial. Still, it might be tempting to conclude that everything fits and that the reason for the hemispheric asymmetry of *maria* is now clear. Alas, such is not, in fact, the case.

Zhong's modeling showed that the hemispheric pattern of convection would only arise if the radius of the lunar core is 250 kilometers or less. Recent measurements suggest that the core radius is likely about 350 kilometers, indicating that some other mechanism must have been at work.

Having an Impact

As part of my doctoral research at Washington University in St. Louis, I examined the preservation of lunar impact basins with the objective of learning about the early thermal history of the Moon. The focus of the study was *viscous relaxation*, a phenomenon that goes on in many places in the solar system.

On Earth, in regions where the crust is thick enough and where temperatures of the lower crust are sufficiently high, the bottom portion of the crust, though still solid, is able to flow plastically over time, acting in many ways like an extremely viscous fluid. These movements are able to even out to some extent variations in the thickness of the crust. A prime example of this process on Earth occurs beneath the Himalayas. There, the Indian sub-continent sits on a tectonic plate that is colliding with Asia, with the result that the crust at the boundary is compressed and thickened. The viscosity of rock decreases with increasing temperature, so at great depths under the Himalayas, the crust is able to flow and relax the stress imposed by all the stuff that's been piled on top of it.

On the Moon, most of the significant variations in topography and crustal thickness are a consequence of large impacts that occurred billions of years ago, leaving obvious gouges in the form of deep basins. You can learn something about the temperature in the lower crust at the time a given basin formed by looking to see whether or to what degree things relaxed afterward. And by putting together the results from all lunar basins, you can get an idea about the evolution of temperatures in the lower crust over time. The trick is to know what the various impact basins looked like when they were first excavated.

The collision of an asteroid or comet with a planet produces results that are similar to a buried explosion. Indeed, they are so similar that in the 1950s buried nuclear explosions were used to study the dynamics of how impact craters arise on Earth and other celestial bodies. At the moment of contact, the collision unleashes a shock wave that travels through both the target planet and the incoming body, releasing a large amount of energy.

At the speeds that rocky objects typically hit the Moon, the projectile will vaporize completely, leaving only a crater on the surface. Small objects produce simple, bowl-shaped craters, but with larger (or more speedy) ones, the initial bowl-shaped crater collapses, resulting in a more complex structure that may involve a central peak or multiple rings. Here the stresses during impact are so great that the rock surrounding the impact site behaves like a fluid for a short time. The result is similar to what happens when you throw a stone into a quiet pond: A depression forms at first, but the surface of the water quickly rebounds, rising above the equilibrium level before it falls back down. These oscillations repeat for a while, creating a series of circular ripples.

In the case of an impact, this process is arrested abruptly once the stress drops sufficiently for strength to return to the shocked rock. This is why small craters often show just a central peak, whereas larger ones contain a central ring, and the largest sport multiple rings.

The lunar impact basins that have most to contribute to the questions at hand are multi-ring basins, which have diameters of at least 400 kilometers. The gravity measurements made over such basins are quite revealing. In particular, gravity tends to be anomalously strong (in the jargon of the trade, these basins are said to show positive gravity anomalies). This observation is somewhat counterintuitive, because these basins are essentially big holes in the ground, and where mass is missing you might expect to see a negative gravity anomaly.

The positive gravity anomalies associated with multi-ring impact basins appear to stem from two principal sources. First, lava flooded into many of the basins on the lunar near side, leaving thick deposits of dense *mare* basalt. Also, the oscillatory deformation that created such large basins often reached down to the base of the crust, uplifting it and bringing dense mantle material close to the surface, where it, too, enhances the tug of gravity.

Time to Relax

The observation of positive gravity anomalies over basins is not, however, universal. Despite the colossal amount of material excavated, the crust beneath the older basins does not appear to be much thinner than it is in surrounding areas. One explanation is that at the time of these impacts, the magma ocean had just solidified, and the lower crust was still sufficiently hot to be able to flow back into the thinned region. Thus, there appears to be a fairly sharp distinction between older basins, which have relaxed in this way, and younger ones, which have preserved their original geometries.

What conditions exactly would be required to allow relaxation? To answer that question, I developed a numerical model of planetary deformation. The model considers the crust and mantle of the Moon to be *viscoelastic solids*: That is, they respond to loads with an instantaneous elastic response, followed by viscous relaxation of the stress over a longer time scale (determined by the physical properties of the material). An example of such viscoelastic behavior closer to home is the rebound of continents following the end of the Ice Age. The weight of the huge ice sheet covering North America pushed the ground down hundreds of meters. When the ice sheet retreated some 10,000 years ago, the ground quickly began to bounce back to its original position. But the motion was not instantaneous; indeed, in many places it continues today. By measuring the rate of this slow rebound, geophysicists have been able to estimate the viscosity of the Earth's nominally solid mantle.

In the case of the Moon, I used the known material properties of lunar rocks and calculated the resulting viscosity variation with depth for a variety of possible temperatures. I then ran my numerical model using the geometry of one well-preserved lunar basin to see how it would evolve over time under different thermal conditions. I found that relaxation occurred much more readily in thicker crust, because the plastic deformation of rock is enhanced when it is deeply buried (and thus hot).

Large impacts act to thin the lunar crust significantly, by about 20 to 40 kilometers, which in turn cools it. As a result, the crustal thinning produced by the largest impacts tends to prevent

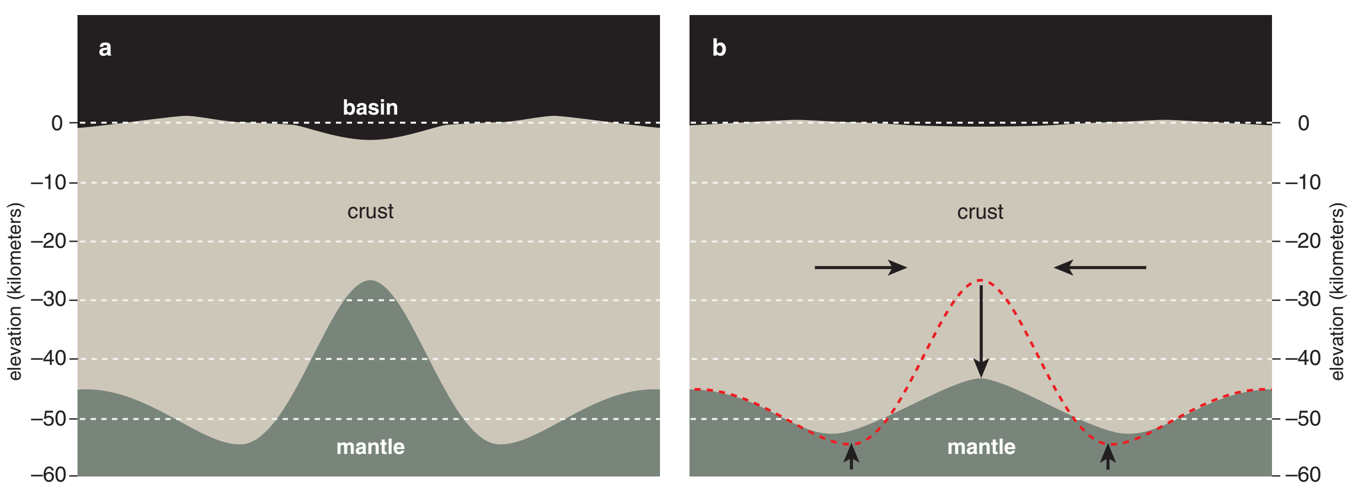


Figure 7. Many of the largest impact basins display positive gravity anomalies because at the time of their formation dense mantle material rose toward the surface *(left)*. But if the crust is sufficiently hot, the deeply buried mantle peak can subside over time as the overlying crust flows inward and thickens *(right)*, leaving no significant gravity anomaly.

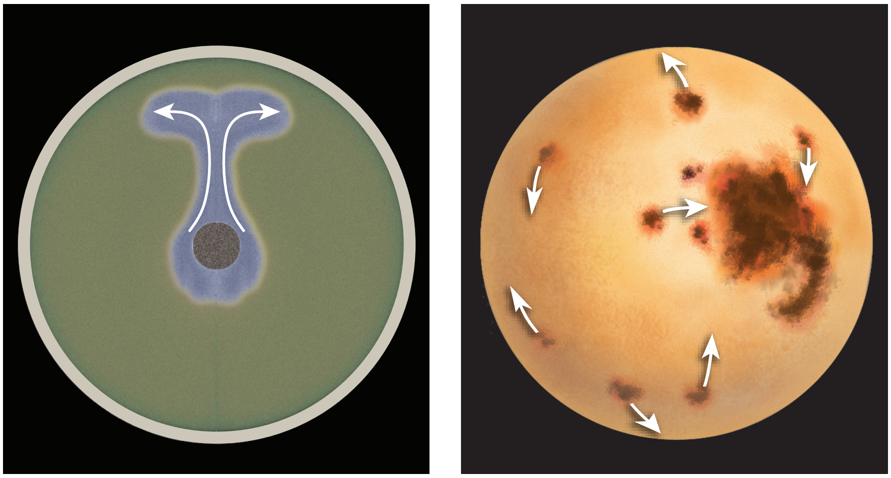


Figure 8. Theories to account for the Moon's hemispheric asymmetries remain tentative. One idea is that rock rich in radioactive elements first sank and then rose from deep within the lunar mantle, creating an upwelling that thinned the crust in one hemisphere and prompted basaltic lavas to fill basins there *(left)*. Another notion is that the hemispheric asymmetries arose much earlier, during the magma-ocean phase. Floating solids—"rockbergs"—which had been moving about randomly, clumped together in one "continental" hemisphere first *(right)*, causing the crust there to become especially thick. This process initially left radioactive elements in the magma; these were taken up in the rocks of the opposite hemisphere after the magma ocean cooled completely.

subsequent relaxation. Indeed, the near-side crust is so thin that an impact forming a basin with a diameter greater than 500 kilometers will remove almost the entirety of the crust, making it nearly impossible to achieve flow in this layer.

By contrast, the thicker far-side crust does allow for viscous flow near the bottom—or rather, it once did. The temperatures required are within about 200 degrees Celsius of those that prevailed in the lower crust just after final solidification of the magma ocean. As a result, viscous relaxation could only have taken place during a relatively short period following the last stages of the magma ocean's existence.

The numerical models I ran showed that even where viscous relaxation evened out uplift of the crust-mantle boundary after an impact, the rugged topography created at the surface would be preserved. The colder the temperature of the crust, the more the initial surface topography was retained.

My modeling did not, however, explain why many impact basins had completely relaxed—particularly some of those found on the near side, where the crust is thin and, at least in theory, should not have been able to relax. One possible solution to this quandary is that these basins formed at a time prior to full solidification of the magma ocean, when the lower crust would have been partially molten. The presence of liquid within the lower crust would greatly facilitate flow, making relaxation possible even for the thin crust of the near side. Alternatively, an energetic impact into such weakened crust may have resulted in complete collapse of the transient craters immediately after they had formed, leaving no topographic basins behind, as is observed on Europa, one of the icy satellites of Jupiter.

In either case, it is interesting to note that the crust of the far side's South Pole-Aitken basin has remained thin, although it is the oldest (and largest) impact basin on the Moon. This observation suggests that whatever thermal conditions prevailed in the lower crust during the formation of the near-side basins didn't apply on the far side.

These hemispheric asymmetries in temperature and crustal thickness may both have resulted from the dynamics of the way the magma ocean solidified. Recall that the lunar crust formed by flotation of light minerals, principally plagioclase feldspar, in the magma ocean. Random fluid motions likely caused plagioclase crystals to coalesce into "rockbergs" above places where the magma was sinking. Haphazard horizontal motions could then have caused these rockbergs to agglomerate into a single huge "continent," thus forming the heart of the far-side crust and producing, ultimately, a distinct hemispheric asymmetry in crustal thickness.

If this surmise is correct, the magma ocean must have solidified first on the far side, concentrating the remaining liquid—enriched in incompatible radioactive elements—on the near side. The radioactive heating these elements provided would have resulted in very slow cooling of the near-side crust. In addition, ilmenite-rich cumulates would have formed primarily on the near side. Thus, this half of the Moon would have possessed both the parent materials for the *mare*-forming lavas and the heat sources necessary to melt these rocks.

This scenario neatly accounts for the Moon's obvious asymmetries: the greater crustal thickness and prevalence of viscous relaxation on one side, along with the concentration of KREEP and *mare* basalts on the other. What the theory doesn't explain, though, is why the two contrasting hemispheres of the Moon are aligned such that you see only one of them when you gaze into the sky at night.

Alignment Job

One reason why the two contrasting sides of the Moon are aligned as they are may be, well, no reason—just simple coincidence. This possibility isn't particularly satisfying, but it may well be the best explanation. Some support for this view comes from a more detailed examination of the Moon's asymmetries.

It turns out that the hemisphere containing the thickest crust is not, in fact, completely congruent with the far side. Rather, the hemisphere with the thickest crust largely overlaps the far side but is canted 23 degrees away from the Earth-Moon axis. In addition, the offset between the Moon's center of mass and the center of this spherical body (an offset created by the anomalously dense *mare* basalts) is canted by about the same amount. Were some gravitational mechanism at work, you'd expect to find the Moon's two contrasting hemispheres better aligned with Earth.

David E. Loper and Christopher L. Werner of Florida State University proposed an intriguing solution to this problem. They posited that the Moon's asymmetries arose by virtue of a phenomenon known as "tilted convection", an idea championed by Ruby Krishnamurti and Louis N. Howard, also at Florida State University.

When a fluid layer is heated from below or cooled from above (or both), its behavior depends strongly on the ratio between the driving force (the density contrast between fluid at the top and bottom of the layer created by the expansion of the hot fluid relative to the cold fluid) and the resisting force (primarily the viscosity of the fluid). Once this ratio rises above a critical value, the hot fluid at the bottom of the layer will begin to rise and the cold fluid at the top will begin to sink, a much more efficient mechanism than conductive cooling.

At first, this motion takes the form of stable "cells", in which the rising and sinking parts of the fluid form a simple geometric pattern. Such motions can be easily observed by ordering a bowl of hot miso soup at a Japanese restaurant and watching the grainy soybean particles as they track the movement of the liquid.

If you turn up the heat on such a liquid, it will eventually begin to convect more vigorously, and the orderly cells will give way to hot and cold plumes rising and falling at random locations. Krishnamurti and Howard discovered that heating the fluid beyond yet another critical value results in plumes rising and falling at an angle, moving horizontally as well as vertically—in other words, tilted convection.

Because the direction of the tilt could be influenced by imposing a lateral temperature contrast, Loper and Werner theorized that tilted convection in the Moon's magma ocean—which, according to their calculations, should have been convecting with sufficient vigor—could have preferentially swept floating anorthosite crystals to the far side. According to their hypothesis, the slight differences in the surface temperature arose because the near side, being bathed in Earthshine when not illuminated by the Sun, was somewhat warmer than the far side.

This explanation requires that early on the Moon's rotation was "locked," with one side facing the Earth at all times. The best guess is that it would take only 10 million years or so for such locking to occur, which is less than the length of time the Moon's magma ocean probably remained liquid. So Loper and Werner's hypothesis is at least plausible, although it is very difficult to test.

Whatever the explanation for the rough alignment of the Moon's two very different hemispheres, the mechanism by which the asymmetry arose in the first place may be close at hand, following at least the outlines of the thinking I've sketched above. But further investigation will be needed if these ideas are to be fully confirmed.

Some additional evidence may come from the armada of missions to the Moon slated to be undertaken over the next five years. The United States, Europe, Japan, China and India all plan to explore the Moon with robot probes, including a possible sample-return mission to South Pole-Aitken basin. Investigators will have to wait and see whether these expeditions shed light on the contentious and critical issue of the Moon's hemispheric asymmetry, but one thing is certain: Much more will soon be known about our puzzling two-faced neighbor in space.

Bibliography

Jolliff, B. L., J. J. Gillis, L. A. Haskin, R. L. Korotev and M. A. Wieczorek. 2000. Major lunar crustal terranes: Surface expressions and crust-mantle origins. *Journal of Geophysical Research* 105(E2):4197–4216.

Krishnamurti, R., and L. N. Howard. 1981. Large-scale flow generation in turbulent convection. *Proceedings of the National Academy of Sciences of the U.S.A.* 78:1981–1985.

Loper, D. E., and C. L. Werner. 2002. On lunar asymmetries: 1. Tilted convection and crustal asymmetry. *Journal of Geophysical Research* 107(E6):5046.

Herbert, F., M. J. Drake and C. P. Sonett. 1978. Geophysical and geochemical evolution of the lunar magma ocean. In *Proceedings of the 9th Lunar and Planetary Science Conference*, ed. R. B. Merrill, 1:249–262. New York: Pergamon Press.

Mohit, P. S., and R. J. Phillips. 2006. Viscoelastic evolution of lunar multiring basins. *Journal of Geophysical Research* 111(E12001).

Schenk, P. M. 2002. Thickness constraints on the icy shells of the Galilean satellites from a comparison of crater shapes. *Nature* 417:419–421.

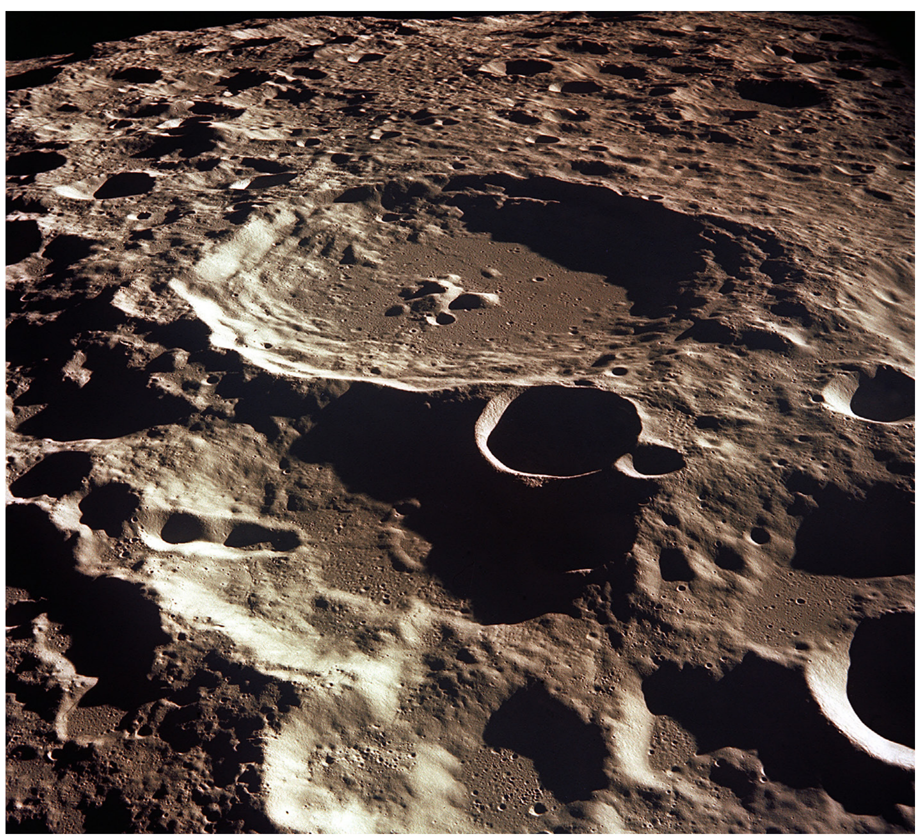
Stegman, D. R., A. M. Jellinek, S. A. Zatman, J. R. Baumgardner and M. A. Richards. 2003. An early lunar core dynamo driven by thermochemical mantle convection. *Nature* 421:143–146.

Wieczorek, M. A., B. L. Jolliff, A. Khan, M. E. Pritchard, B. P. Weiss, J. G. Williams, L. L. Hood, K. Righter, C. R. Neal, C. K. Shearer, I. S. McCallum, S. Tompkins, B. R. Hawke, C. Peterson, J. J. Gillis and B. Bussey. 2006. The constitution and structure of the lunar interior. *Reviews of Mineralogy and Geochemistry* 60:221–364.

Zhong, S., E. M. Parmentier and M. T. Zuber. 2000. A dynamic origin for the global asymmetry of lunar mare basalts. *Earth and Planetary Science Letters* 177:131–140.

For relevant Web links, consult this issue of *American Scientist Online:*

http://www.americanscientist.org/Issue TOC/issue/1081



NASA

Phoenix on Mars

The latest successful landing craft has made new discoveries about water on the red planet

Walter Goetz

Since we received our first close-up photographs of Mars, when Mariner 4 flew by it in 1965, our nearest neighbor has appeared to be much like our own planet in many ways, but also distinctly different. Mars is about half the size and has about 40 percent of the gravity of Earth, it's at least 55 million kilometers away (depending on the two planets' positions in their orbits), and it currently takes at least nine months to get there. But like Earth, Mars has polar ice caps, clouds in its atmosphere and seasonal weather patterns. It has familiar geological features, such as volcanoes and canyons. However, although there are signs of floods in the ancient past, Mars is now apparently a barren world.

What is the history of liquid water on Mars? Has water ever been stable on its surface (or in its near subsurface) for a geologically significant period of time? Was Mars warm and wet in ancient times? If so, what triggered the apparent change in climate? And could primitive terrestrial life-forms evolve in the present or past Martian environment? These are the main questions that have driven the exploration of Mars since the mid-1960s. In addition, if humans ever tried to travel to, or even set up an outpost on, another planet, Mars would likely be the first choice, so there's even more reason to learn as much as possible about our neighboring planet.

Missions to Mars have been a mix of failure and success. The first working spacecraft to land on the planet's surface were Viking 1 and 2 in the mid-1970s, and they returned the first color images of the planet. They also sent back data long past their planned mission lifetime, until 1982 and 1980, respectively. Their experiments on Martian soil, looking for signs of microscopic life, were inconclusive. More than a decade later, a mission to send an orbiter to Mars ended in failure, but another, Mars Global Surveyor, arrived in 1997 and returned data until October 2006. Also in 1997, the Mars Pathfinder lander, with its Sojourner rover, landed safely and was remarkably successful.

In 1999 the Mars Climate Orbiter and the Mars Polar Lander both failed and were lost upon arrival at Mars. The Mars Surveyor 2001 mission, including an orbiter, a lander and a rover, was canceled in 2000, but its orbiter was repurposed and successfully launched as the 2001 Mars Odyssey orbiter. This orbiter has also relayed information back from the twin rovers, Spirit and Opportunity, which landed in 2004. The European Space Agency (ESA) saw the safe arrival of its orbiter, Mars Express, in 2003, although the lander was lost on deployment. The Mars Reconnaissance Orbiter safely joined Mars's orbit in 2006, providing the highest camera resolution yet.

However, after the Mars Polar Lander crashed and the Mars Surveyor 2001 mission was canceled in 2000, there seemed to be no hope for a new mission to the Martian arctic regions. The situation changed early in 2002 when the Mars Odyssey orbiter discovered large amounts of near-surface hydrogen in exactly these regions. The hydrogen reservoir was interpreted as water ice—less than a meter below the surface. It was argued that such arctic water ice might contain the long-searched for (and long-missed) organic compounds that could signify the presence of life, either past or present.

These discoveries led a group, headed by Peter H. Smith of the University of Arizona, to develop a mission that would build on previous designs and use the already completed, but unused, Mars Surveyor lander. Thus was born the Phoenix Mars Lander, named because like the mythical bird, it had been resurrected from the ashes of its predecessors. The rocket that carried Phoenix was launched on August 4, 2007, and the spacecraft landed safely on May 25, 2008.

Anatomy of a Lander

Phoenix's suite of scientific instruments includes several imaging systems that have different levels of resolution. From lowest to highest resolution, these instruments are its stereo surface imager (SSI), which can show about 1 millimeter per pixel; a robotic arm camera (RAC), with a resolution of more than 24 micrometers per pixel; an optical microscope that can reach about 4 micrometers per pixel; and an atomic force microscope (AFM) that can show about 0.1 micrometers per scan. Our group at the Max Planck Institute for Solar System Research, in collaboration with the University of Arizona, contributed the RAC and the focal-plane assembly of the optical microscope.

The lander also has a wet chemistry laboratory unit (WCL), where it can mix Martian soil in liquid water. The unit consists of four such cells, each designed for a single use. The resulting aqueous solution is analyzed by ion-selective electrodes, which provide information on the compounds in the soil (such as salts) that are soluble in liquid water.

Another important instrument is a thermal analyzer, designed to heat soil

Walter Goetz received his Ph.D. from the Center for Planetary Science at the University of Copenhagen in 2002. He is a physicist at the Max Planck Institute for Solar System Research in Katlenberg-Lindau, Germany, and a member of the team that developed the robotic-arm camera for the Phoenix lander. He participated full time with the Phoenix primary mission operations at the University of Arizona, Tucson, in 2008, and with the Mars Exploration Rover primary mission operations at the Jet Propulsion Laboratory in 2004. He is part of the team developing instruments for the European Space Agency's upcoming ExoMars rover. Address: Max Planck Institute for Solar System Research (MPS), Max-Planck-Str. 2, 37191 Katlenburg-Lindau, Germany. Email: goetz@mps.mpg.de

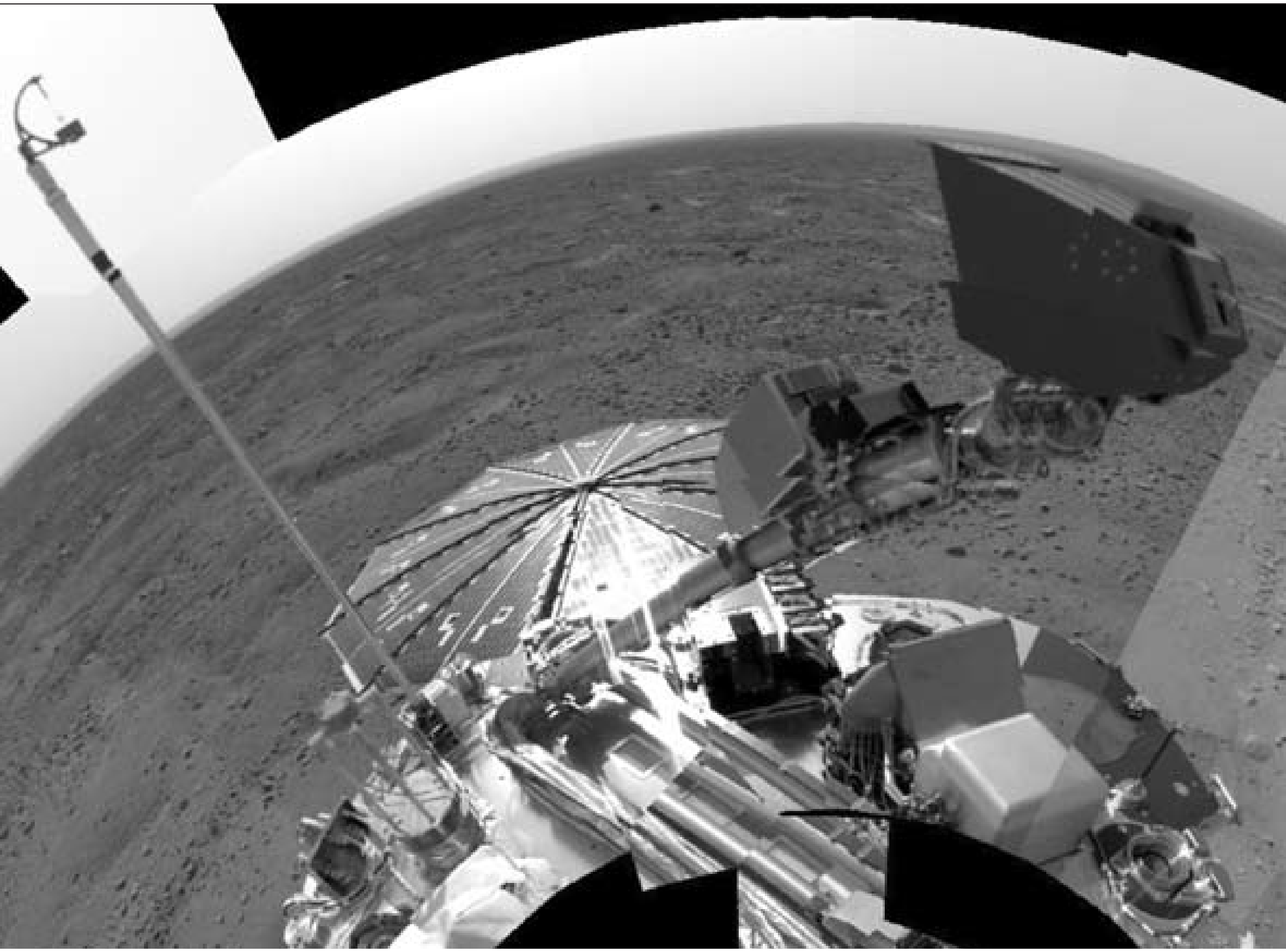


Figure 1. More than a simple tourist snapshot of itself in an exotic locale, this self-portrait of the Phoenix lander shows a remarkable feat: the safe arrival, descent and landing of the spacecraft onto the surface of Mars, a result that is still not a given for interplanetary missions. Not only that, the lander also carried out a series of studies on water ice, soil chemistry and weather patterns over the course of several months in 2008. Phoenix is a Cinderella story, as the lander was originally constructed for the cancelled 2001 Mars Surveyor mission. Its resurrection and deployment to Mars' polar regions has provided great insights into the water cycle on our neighboring planet. This panorama of the lander, showing its robot arm partially deployed, was taken during the first few days after landing. (Image courtesy of NASA/JPL/University of Arizona and M. T. Lemmon, Texas A&M University.)

samples up to 1,000 degrees Celsius; the evaporating gases are then studied by mass spectroscopy. This instrument group (referred to as the Thermal and Evolved Gas Analyzer, or TEGA) should be able to characterize the inventory of potential organic compounds in the Martian soil by detecting either the parent organic molecules or their thermally generated fragments, as the temperatures where specific gases are released constrain the identity of the parent compound. As the amount of heat is increased at known levels, any additional increase in temperature also reveals phase transitions in the compounds, which can possibly be identified by their enthalpic characteristics.

The spacecraft also has a robotic arm that is in itself a scientific instrument, as it allows the lander to characterize the physical properties of the soil. The scoop at the end of the 2.3-meter-long arm allows controllers to select and transfer specific soil samples to various instruments. An ice drill is mounted to the backside of the scoop. The robotic arm camera is positioned on the arm so that it can see into the scoop and image the collected soil sample at high resolution. A sensor mounted next to the scoop can measure the soil's electric and thermal conductivity between four needles. An additional sensor can measure the atmospheric water vapor pressure and the relative humidity of the atmosphere.

A meteorological mast, provided by the Canadian Space Agency, collects data about the Martian weather that help describe how water cycles between the solid and gas phases at the landing site. Its central instrument is a LIDAR (Light Detection And Ranging) that probes the vertical structure of the atmosphere by measuring the travel time of its emitted light as it is backscattered by suspended particles (such as dust and ice) in the air. Pressure and temperature sensors are

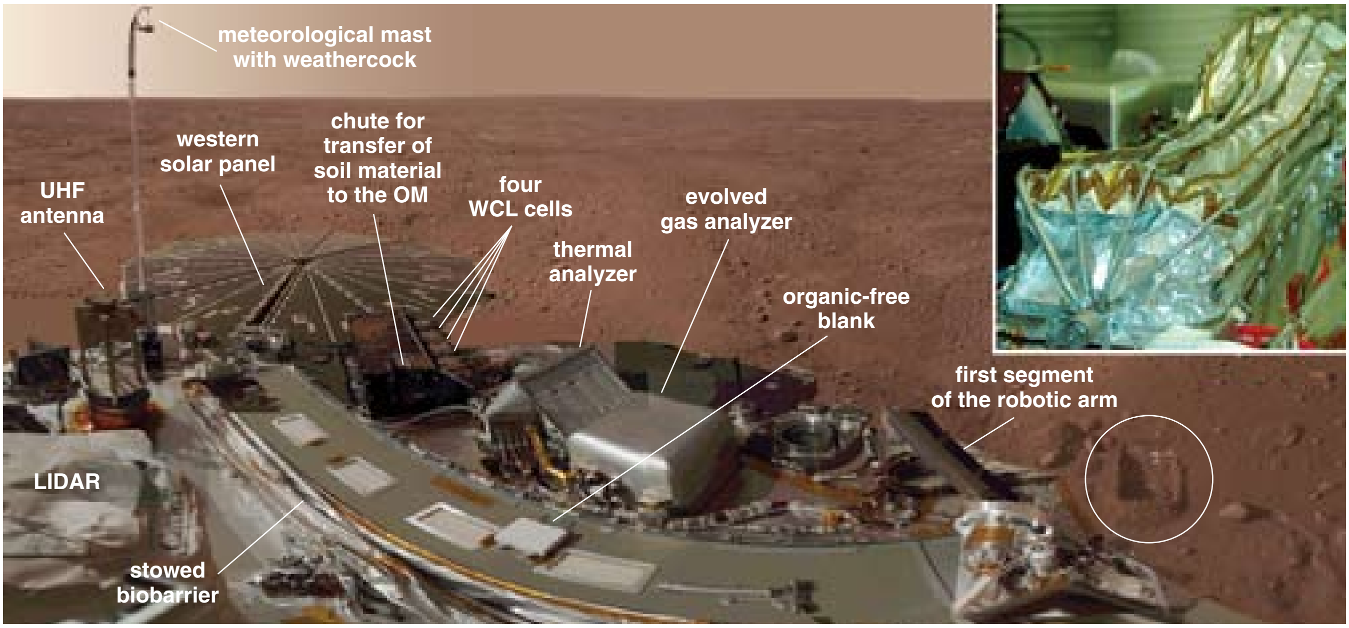


Figure 2. An image of the lander by its stereo surface imager shows a number of Phoenix's scientific instruments. At left, a LIDAR (Laser Detection And Ranging) is used in conjunction with the metorological mast and weathercock for weather studies. A UHF (ultra-high frequency) antenna relays data to an orbiting satellite. The lander has two large circular solar panels, the western-facing one of which is shown. A chute is used to transfer soil samples to the optical microscope (OM). A wet chemistry laboratory (WCL) has four cells in which it can mix soil samples with liquid for analysis. A thermal analyzer and evolved gas analyzer heat samples and look at the gases emitted, respectively. A Teflon block, called an organic-free blank, is used to serve as a baseline reference for organic molecules. One segment of the robotic arm can be seen at the right of the image, next to a trench about 20 centimeters wide *(white circle)* that was later analyzed for water ice. The inset at upper right shows a foldable cover, called a biobarrier, that protected the robotic arm and scoop from biological contamination during flight. This cover folded towards the front of the lander and was stowed after landing. (Image courtesy of NASA/JPL/University of Arizona and M. T. Lemmon, Texas A&M University; inset image courtesy of IEEE and R. G. Bonitz, NASA/JPL.)

mounted to the mast at three different heights above the deck (25, 50 and 100 centimeters), and the mast is topped by a Danish-produced weathercock (telltale, or wind indicator). The SSI provides complementary data on the atmospheric dust opacity and water vapor abundance by imaging the solar disk through specific visible and near-infrared filters.

Where It All Happens

Spectroscopic data and high-resolution images from various orbiters (including Mars Global Surveyor, Odyssey, Mars Express and Mars Reconnaissance Orbiter) were available prior to landing and were used extensively to select the best landing site for Phoenix, both in terms of safety and for the best chance of doing useful research.

The distance from the landing site to the northern border of the volcanic Tharsis region is about 500 kilometers and to the nearest volcano (Alba Patera) is about 1,800 kilometers. The north-polar ice cap and the circumpolar dunes are located about 2,000 kilometers north of the lander. On a large scale, we expected volcanic ashes from the Tharsis province as well as sand grains from the north-polar dunes at the site. The landing site is also situated about 20 kilometers west from the Heimdall crater, which has a diameter of 11 kilometers and a depth of about 1 kilometer. Thus ejected soil from these depths may also be found at the landing site.

A few days after landing, the terrain below the spacecraft was examined by the RAC in order to confirm the stability of the spacecraft's position. The first image showed a bright, even surface (called "Holy Cow"; the nomenclature followed fairy-tale themes) that was uncovered by the action of the descent thrusters. Apparently, the subsurface ice discovered by Odyssey in 2002 was right there—only a few centimeters below the surface. The images suggest that this is ice-rich regolith rather than pure water ice. Over the course of about 50 Martian days (or sols), another icy soil patch, dubbed "Snow Queen" and located just next to Holy Cow, developed numerous cracks after it lost its thermally insulating blanket of soil.

During the Phoenix mission, 12 trenches were excavated, the deepest being 18.3 centimeters. The appearance of the subsurface soil was different from trench to trench. In some trenches (such as one dubbed "Dodo Goldilock") almost pure ice was found, as determined by the spectra acquired by the SSI. Other trenches yielded ice-rich regolith, whereas in some no ice was found at all. In the Dodo Goldilock trench, bright centimeter-sized clumps disappeared over the course of four sols. This observation suggests that the bright material in the shallow subsurface is indeed water ice. So far, it is not clear why the ice/regolith mixing ratio varies so much within a few meters.

The surface of the Martian polar environment takes on a hummocky appearance that might account for some of the soil inconsistencies. The basic model for the formation of these "polygons" was developed by Ronald Sletten and his colleagues at the University of Washington. Seasonal contraction and expansion of soil generates wedge-shaped fractures. During winter, fine-grained debris moves into these wedges and prevents them from completely closing again during the next summer. The seasonal stress generated by these processes is relaxed by the formation of mounds (or polygons) at a certain spatial frequency. The net result is a slow cyclic transport of soil material. This erosional process is known as cryoturbation and occurs frequently in terrestrial environments around the edges of glacial regions.

The Heimdall crater formed about 500 million years ago. It seems likely that

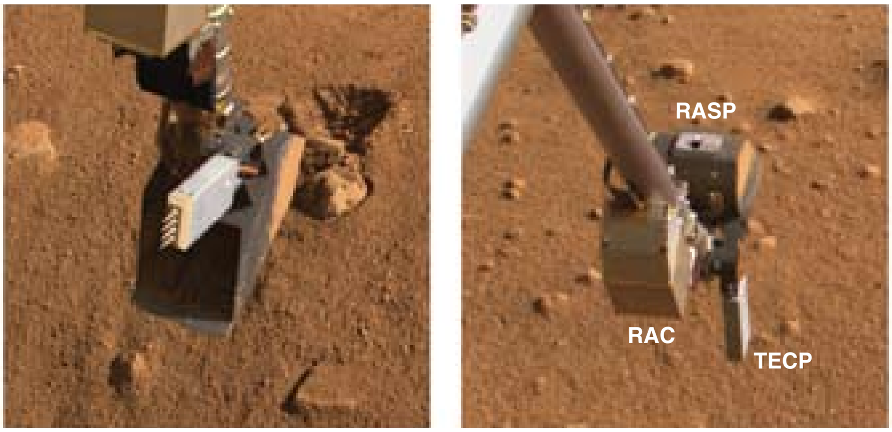


Figure 3. Images from Phoenix's stereo surface imager show the end of the lander's robotic arm as it digs a sample of Martian soil *(left)*. As the arm lifts up *(right)*, its instruments can be seen: an ice drill (called the Rapid Active Sampling Package or RASP) mounted on the back of the scoop, the robotic arm camera (RAC) positioned so it can see into the scoop and image soil samples and a soil sensor (called the Thermal and Electrical Conductivity Probe, or TECP) mounted next to the scoop. (Images courtesy of NASA/JPL/University of Arizona and M. T. Lemmon, Texas A&M University.)

excavation of this crater contributed soil material to the landing site. Cryoturbation processes, however, take place over a much shorter time scale, continuously renewing the landscape and making the Phoenix landing site the youngest among all other past Martian landing sites (those of Viking, Mars Pathfinder or the Mars Exploration Rovers).

Larger rocks or boulders (bigger than 20 centimeters or so) are absent at the landing site. Water ice is abundant near the surface, in agreement with Odyssey 2002 data. Perhaps the absence of larger rocks can be explained by the high concentration of condensed volatiles (such as water ice) in the subsurface that were affected by the Heimdall impact: A violent explosion would have removed and crushed the rocks that may have been at the landing site initially. A future, systematic study of the correlation between rock density and distance to the nearest crater may provide further understanding of the size distribution of rocks at the site.

The Soil Itself

Microscopic color images from Phoenix demonstrate the great diversity of particles in Martian soil. AFM scans have given three-dimensional representations of dust particles, but it is unclear how typical these particles may be of Martian dust in general. Reddish-orange dust dominates by volume. The individual dust particles cannot be resolved by the microscope and must therefore be on the order of 10 micrometers or less in size. According to a preliminary classification, two different types of grains are present in the soil: Reddish-brownish to colorless grains and dark (almost black) grains. The origin of these grains is uncertain, but a careful comparison to terrestrial analog soils may constrain the potential scenarios for the formation of these grains.

Key Phoenix instruments, such as TEGA and WCL, have provided new insights into the microscopic structure, as well as the mineralogy, of the soil. One cubic centimeter of material transferred to one of the WCL cells was mixed with 25 cubic centimeters of aqueous solution and produced a weak alkaline solution (with a pH of about 8.3) that contained

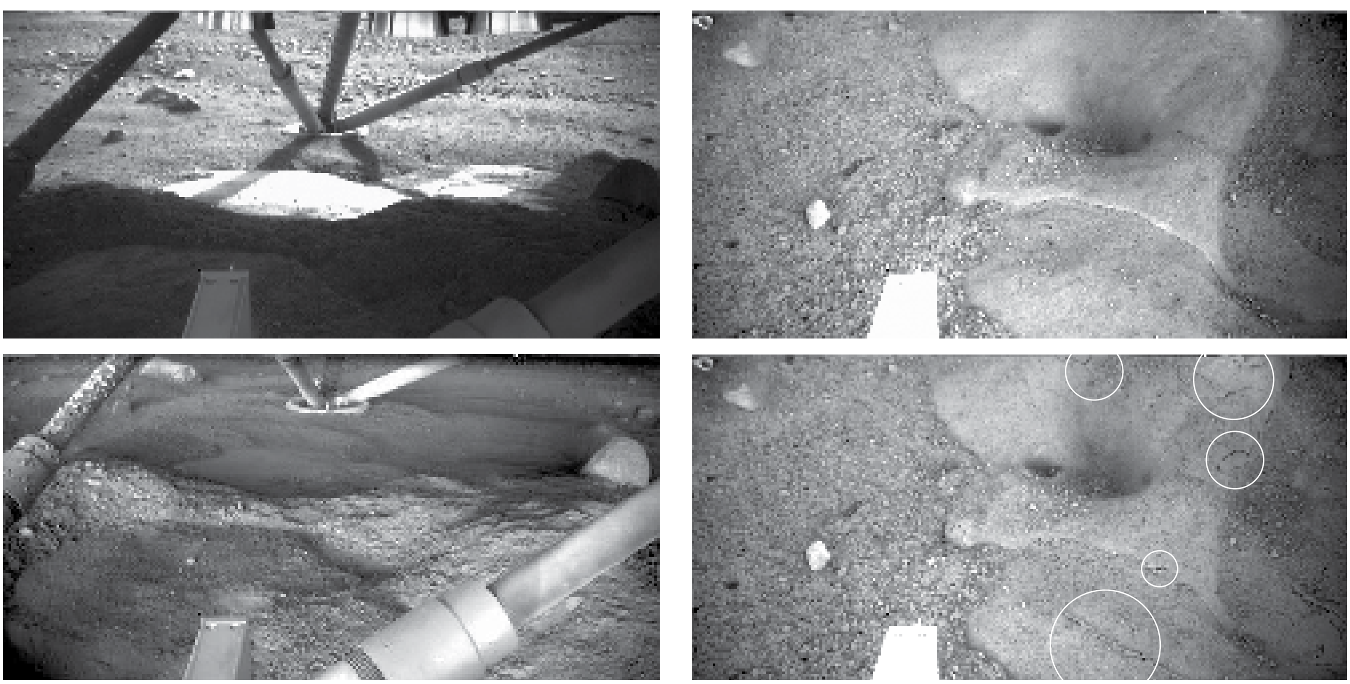


Figure 4. The icy area dubbed "Holy Cow" was swept clean by the lander's descent thrusters. It was imaged by the robotic arm camera, the only onboard instrument able to see beneath the lander; the light rectangle in all four images is the arm's soil sensor. In sunlight, Holy Cow appears bright and reflective *(top left)*, but during the reddish light of twilight, the patch is about as bright as the surrounding soil, indicating that it is not pure ice *(bottom left)*. A neighboring region, "Snow Queen," was also uncovered by the lander's thrusters. The region was initially smooth *(top right)* but showed surface fractures after about 50 Martian days *(bottom right, in white circles)*, indicating the sublimation of water ice. (Images courtesy of NASA/JPL/Univeristy of Arizona, and H. U. Keller and W. J. Markiewicz, Max Planck Institute for Solar System Research, Katlenburg-Lindau.)

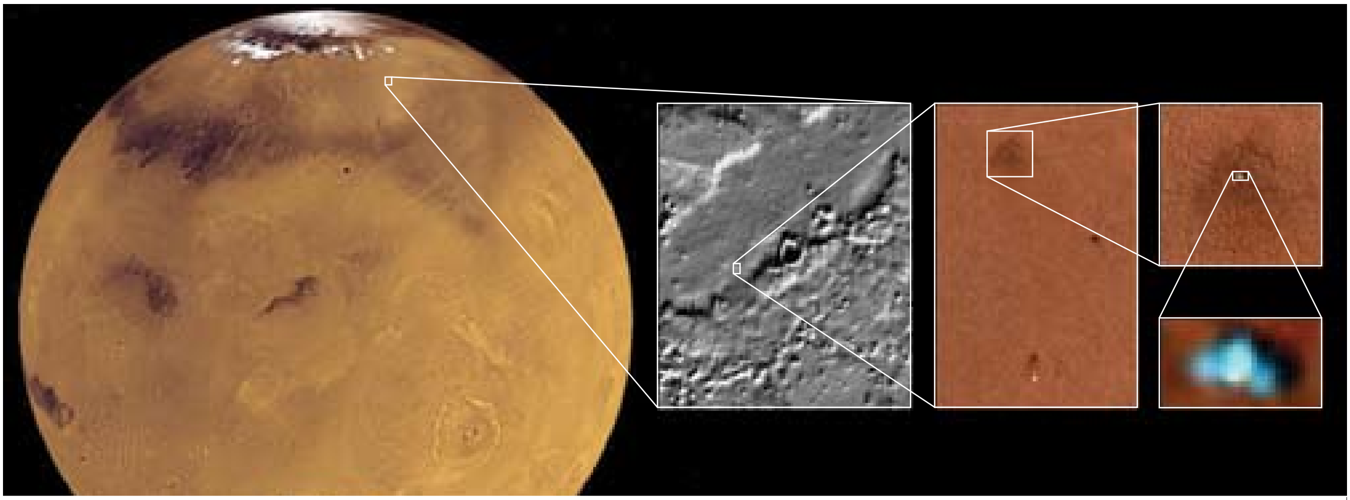


Figure 5. From global to local scale, successively enlarged images zoom in on the Phoenix lander on the Martian surface. The landing site is near the northern polar ice cap of Mars *(far left)*. A black-and-white image about 280 meters wide from Mars Global Surveyor shows that the landing site is off to the left of the circular Heimdall crater *(second from left)*. A higher resolution orbital image, taken 22 hours after landing by the HiRISE camera aboard the Mars Reconnaissance Orbiter, barely shows the lander; the black dot at the middle right of the image is the ejected heat shield, and the bright dot near the bottom center is the parachute *(middle)*. An enlargement of the middle image *(top right)* shows surface roughness and coarser-grained, darker material that was exposed by the descent thrusters. A final enlargement *(bottom right)* shows the lander deck and two solar panels, at a resolution of about 33 centimeters per pixel. (Images courtesy of NASA/JPL, Malin Space Science Systems and the University of Arizona.)

surprisingly large quantities of perchlorate (ClO_4^-), salts that could lower the freezing point of water and that have the potential to be found in a liquid-water solution under the temperature and pressure conditions on present-day Mars. This ion was by far the dominant anion (or negative ion) in the solution. Among the cations (positive ions) were, in order of decreasing concentration, magnesium, sodium, calcium and potassium.

To extrapolate from these results, one gram of Martian soil might have a perchlorate abundance of about 1 percent by weight. Such a concentration exceeds that found in some terrestrial desert soils by orders of magnitude. Finding chlorine at the highest possible degree of oxidation has significant implications for our understanding of the chemical processes taking place on the Martian surface, as well as in the atmosphere, and raises several important questions: Is the perchlorate just an exotic compound at the Phoenix landing site, or is it widespread on the surface of the planet? Is the chlorine identified by all previous Mars lander missions mostly present as perchlorate? Even the old question of life on Mars must be reformulated: Which types of primitive (terrestrial) life-forms could have evolved in the Martian soil, given the measured perchlorate concentration?

The ion-selective electrode in the WCL unit that is sensitive to perchlorate, and much less so to nitrate, provided such a strong signal that the identification of perchlorate was unambiguous (the mass of nitrate needed to explain the signal would have exceeded the total mass of the analyzed soil sample). Overall, the suite of electrodes used in

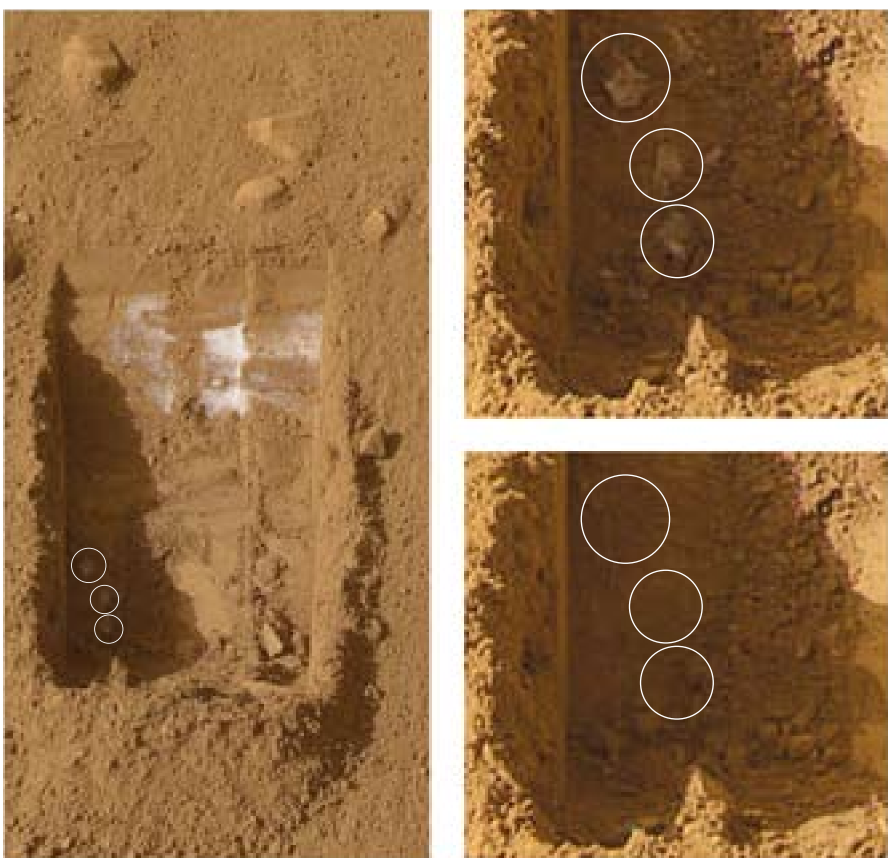
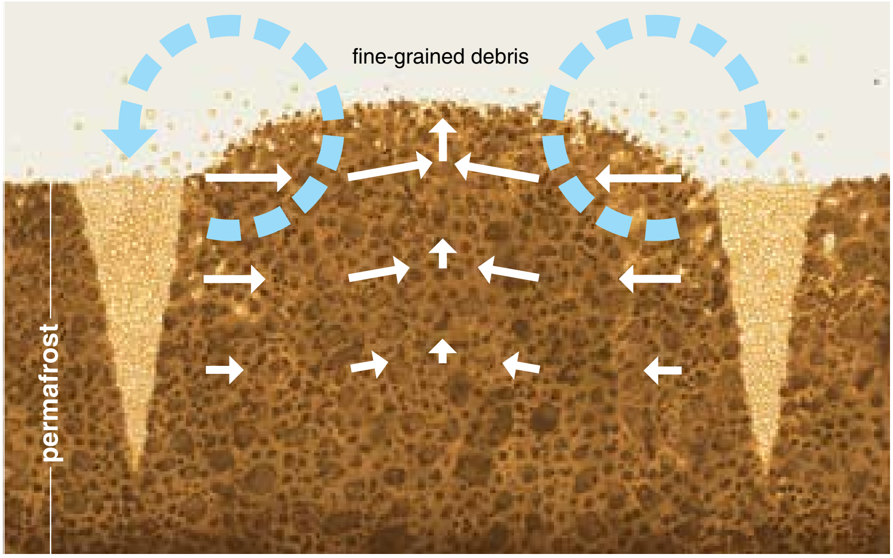


Figure 6. The trench dubbed "Dodo Goldilock" is about 20 centimeters wide. The upper part of the trench reveals nearly pure ice *(left)*. Small clusters of ice particles, about 2 centimeters in diameter, at the lower left of the trench *(white circles)* were visible when the trench was first dug *(see enlargement at top right)*, but by four Martian days later, these spots disappeared *(bottom right)* indicating the particles had sublimated. (Images courtesy of NASA/JPL/University of Arizona and M. T. Lemmon, Texas A&M University.)



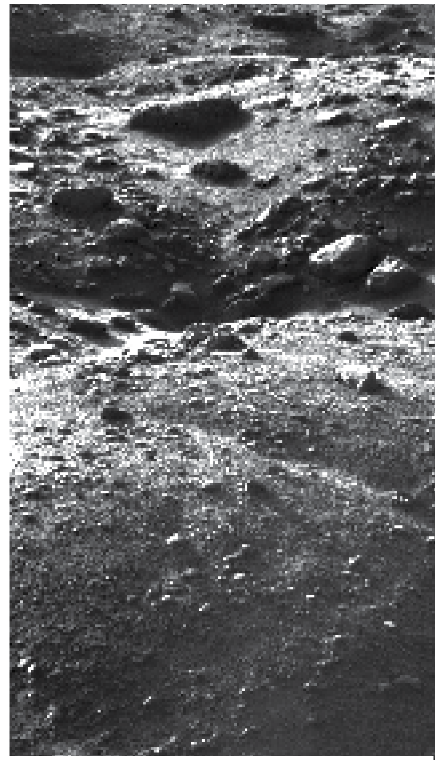


Figure 7. The Martian surface near the polar ice cap often consists of uneven mounds *(right)*, created by a process called cryoturbation. Wedge-shaped gaps form in the soil during the winter *(above)* and partially fill with fine-grained debris that prevents the gaps from fully closing during the summer. The resulting surface stress causes the motion of soil inwards and upwards *(white arrows)* creating these "polygon" surface mounds. Below the level of the permafrost and the depth of the sand wedges, the soil is not perturbed. The blue circular arrows illustrate the long-term transport of soil. (Image courtesy of NASA/JPL/University of Arizona and M. T. Lemmon, Texas A&M University.)

the WCL provided rich information on the water-soluble components of the soil, but each one is sensitive to a range of different ions at strongly different rates. The conversion of the electrode data into concentrations is therefore non-unique, and constraints from other instruments are needed.

Thermal decomposition products of perchlorates were generally not detected by the thermal analyzer or the evaporated gas analyzer's mass spectra, although this does not jeopardize the WCL identification of perchlorate. In at least one of the samples analyzed by these instruments (from a site called "Baby Bear"), some oxygen release in molecular form was observed that may be due to the decomposition of perchlorate.

Another important finding from the mass-spectroscopy studies was the release of carbon dioxide at temperatures of 800 to 900 degrees, which indicates the presence of 3 to 5 percent by weight of calcium carbonate in the soil. The result is remarkable, given that we have been on the hunt for this mineral for many years, and Phoenix found it in the soil! Carbonates are generally the products of aqueous processes, and thus their presence may be indicative of liquid water on the surface of Mars at some point in the planet's history. The inferred presence of carbonates is also compatible with WCL results and explains the alkaline pH of the aqueous solutions.

Furthermore, the absence of certain gases after heating can provide critical information on the mineralogy of the Martian soil: No sulfur dioxide has been released over the entire temperature range (from below 0 degrees, up to 1,000 degrees). This is surprising as all previous lander missions have identified substantial quantities of sulfur in the Martian soil (5 to 10 percent by weight of sulfur trioxide). The presence of sulfate ions is compatible with WCL data. Magnesium sulfate would release sulfur dioxide at temperatures below 1,000 degrees, so the absence of this gas therefore proves the absence of magnesium sulfate in the soil. In the Martian environment (with an atmospheric pressure of roughly 10 millibars, or about a hundredth of that on Earth), calcium sulfate would decompose at about 1,400 degrees, but such temperatures are not reached by Phoenix's thermal analzyer. All these facts taken together point toward the likely presence of calcium carbonate in the soils that Phoenix has analyzed. In fact, large deposits of calcium carbonate previously have been found on the surface of Mars, in particular near the north-polar ice cap.

Nice Weather

Phoenix's instruments have enabled new types of meteorological measurements at the landing site. The site has the advantage that the polar regions exhibit strong weather phenomena, especially cloud formation (as was known from orbital imagery). Also, the weathercock has returned data on wind velocity and direction throughout the mission, enabling fruitful modeling. Phoenix weather measurements were coordinated with orbital observations on a regular basis throughout the mission, strengthening the results.

Mars's atmospheric water vapor pressure, as measured by Phoenix's humidity sensor, rises between about 2 AM and 10 AM, then reaches a plateau (about 1.8 Pascals) that is maintained throughout most of the day. In contrast, atmospheric temperatures continue to rise until about 2 PM. Apparently atmospheric convection becomes very efficient and rapidly redistributes the newly formed water vapor after 10 AM. The spacecraft observed several passing dust devils at typical wind velocities (5 to 10 meters per second). Analysis of the pressure data acquired throughout the mission shows that such dust devils are correlated with brief pressure dips of 1 to 3 Pascals.

LIDAR data from the later part of the mission turned out to be particularly important: Ground fog as well as water-ice clouds near the top of the atmospheric boundary layer (at an altitude of about 4 kilometers) formed every night after sol 80. Many of these clouds had "fall streaks" formed by initially growing, free-falling, then eventually sublimating ice crystals. Such fall streaks also can be observed in terrestrial clouds. Daytime LIDAR data showed mostly dust in the atmospheric boundary layer. However, SSI also documented many daytime clouds late in the mission. In some cases these clouds disappeared by sublimation over a timescale of 10 minutes.

Phoenix's instruments monitored the complete diurnal water cycle: During morning hours water vapor is released into the atmosphere. The sources for the water vapor include the shallow subsurface water ice, water adsorbed to soil grains and, possibly, crystal water in perchlorates. During the night, water vapor condenses and falls out by gravity. Most of these ice crystals sublimate again on their descent through the atmospheric boundary layer. In some cases snowfall was observed, when the fall streaks extended all the way down to the surface.

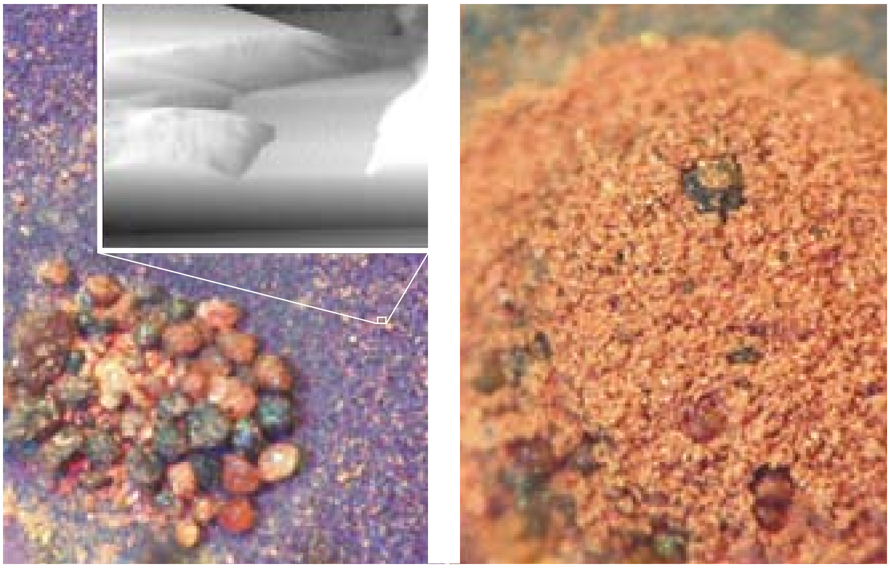


Figure 8. Images of Martian soil from Phoenix's optical microscope show a mix of lighter and darker colored grains, about 60 micrometers in diameter *(left)*, distributed through a matrix of reddish-orange dust, about 10 micrometers in diameter *(right)*. A detail of a dust particle from the atomic force microscope *(upper left)* corresponds to an area about the size of the small rectangle. (Images courtesy of NASA/JPL/University of Arizona, M. H. Hecht, JPL, and W. T. Pike, Imperial College, London; inset courtesy of NASA/JPL and U. Staufer, Technical University Delft, The Netherlands.)

Where Phoenix Is Now

Phoenix surface operations lasted from Martian late spring to late summer—May 26 to November 2, 2008, or 152 sols. The polar night at the landing site lasted from April 1 to July 10, 2009. Since that time, the Sun has again risen above the horizon at the landing site. If the spacecraft—contrary to all expectations—survived both the low temperatures (150 kelvins) of the Martian winter and the dry-ice load built up on its solar panels, it will be able to reanimate itself through a so-called "Lazarus mode." Mars Odyssey was scheduled to search for Phoenix signals starting at the end of 2009.

Independent of its potential reanimation, Phoenix was a highly successful mission that provided on-site geochemical and atmospheric data for the first Martian arctic landing site ever explored. No organic molecules, and no traces of previous or present biological activity, were found at the landing site. Hence, the search for organic molecules will have to be continued by future missions.

It should be noted that organic molecules ought to be present in the Martian soil because of the steady influx of certain types of meteorites that contain

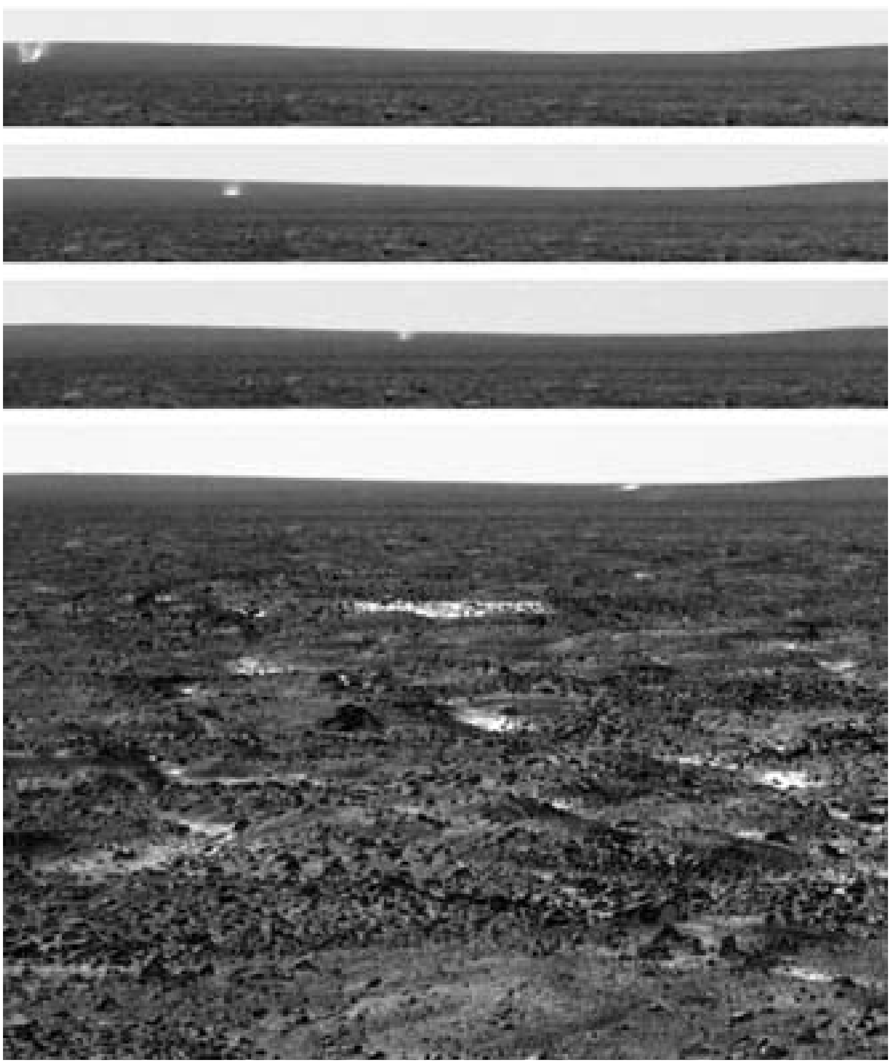


Figure 9. On day 104 of Phoenix's mission, the lander spotted a dust devil about a kilometer away *(top left)*. The dust devil movedto the right and away from the lander *(two middle images, at left)*, ending up about 1.7 kilometers away *(bottom left)*. The presence of the dust devils correlates with brief dips in the atmospheric pressure. Larger dust devils have been observed previously in Mars's Gusev crater. (Images courtesy of NASA/JPL/University of Arizona and M. T. Lemmon, Texas A&M University.)

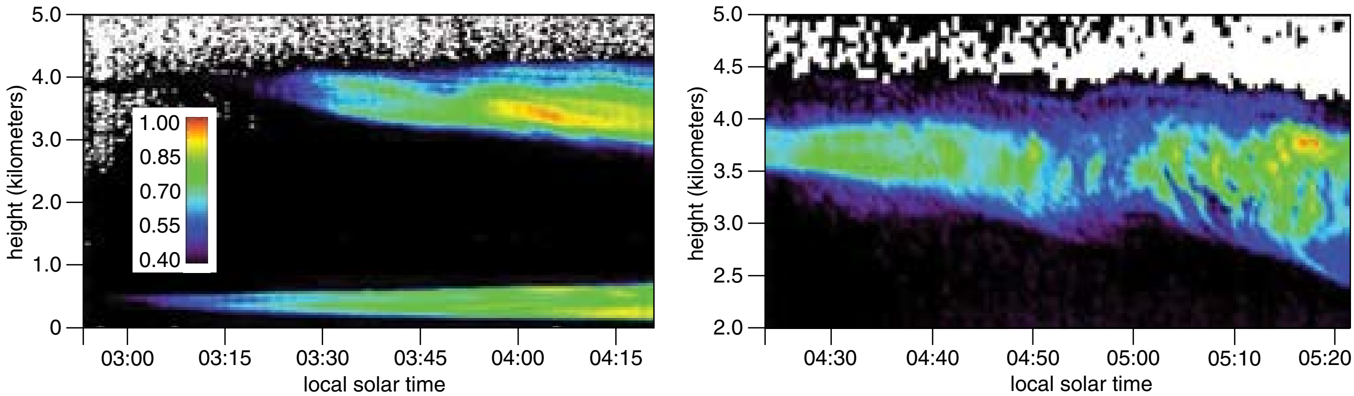


Figure 10. Phoenix's LIDAR was able to characterize the clouds and ground fog that form at night on Mars, as shown by the intensity of light backscattered from ice and other particles *(above left).* Ice clouds at the atmospheric boundary layer, about four kilometers in altitude, often show fall streaks from the sublimation of freely falling ice crystals *(above right).* The size of the ice particles can be calculated from their descent velocities. Similar cloud formations can also be seen on Earth *(right).* (Graphs courtesy of NASA/JPL/University of Arizona and J. A. Whiteway, York University, Canada; photograph courtesy of Marc Thiessenhusen.)

substantial quantities of organic material. The fact that no such molecules have been found in the soil around Phoenix is indicative of fast geological degradation processes. The ever-continuing turnover of soil material (by cryoturbation) at the landing site may have favored such degradation processes.

If organic molecules are ever detected it will be a major scientific task to track down their origin: Are they imported by comets or meteorites, or do they truly attest to primitive, extinct life-forms on the surface of Mars?

Although no organics have been found so far, it is essential to continue this exploration program and search for organic material in more protected environments, such as the interior of sedimentary rocks or deeper soil layers. The next scheduled missions that are available for this task are the rovers Curiosity (which NASA plans to launch in 2011) and ExoMars (which ESA plans to launch in 2018). They will carry complex follow-up instruments that will search for organic molecules in specific equatorial regions. One instrument for ExoMars, the Mars Organic Molecule Analyzer, is presently under development at the Max Planck Institute for Solar System Research.

Curiosity's Sample Analysis at Mars (SAM) instrument has been built by the NASA Goddard Space Flight Center and is ready to go. It will be able to detect particularly low concentrations of potentially organic material. Special attention will be given to the molecule methane (CH_4) whose presence on Mars recently has been demonstrated. The instrument will be able to measure at extremely low methane concentrations (less than one part per billion) the ratio of the isotopes of carbon-13 and carbon-12 present in a molecule. Such data may shed light on the molecule's origin, favoring either geochemical or biological formation.

There is little doubt that Mars will continue to present fascinating new data, surprises and mysteries for these upcoming missions, and to ones still on the drawing board. The two new rovers, Curiosity and ExoMars, will be important benchmarks on that path. Further in the future, robotic return of samples will play a major role in the Mars exploration program. The biggest challenge—a manned mission to Mars—may belong to the distant future, but perhaps at some point such projects will be within reasonable budgets for the major space agencies.

Bibliography

Boynton, W. V., et al. 2009. Evidence for calcium carbonate at the Mars Phoenix landing site. *Science* 325(5936):61–64.

Hecht, M. H., et al. 2009. Detection of perchlorate and the soluble chemistry of Martian soil at the Phoenix lander site. *Science* 325(5936):64–67.

Langevin, Y., F. Poulet, J.-P. Bibring, B. Schmitt, S. Douté and B. Gondet. 2005. Summer evolution of the north polar cap of Mars as observed by OMEGA/Mars Express. *Science* 307:1581–1584.

Sletten, R. S., B. Hallet and R. C. Fletcher. 2003. Resurfacing time of terrestrial surfaces by the formation and maturation of polygonal patterned ground. *Journal of Geophysical Research* 108(E4):8044.

Smith, P. H., et al. 2009. H_2O at the Phoenix landing site. *Science* 325(5936):58–61.

Whiteway, J. A., et al. 2009. Mars water-ice clouds and precipitation. *Science* 325(5936):68–70.

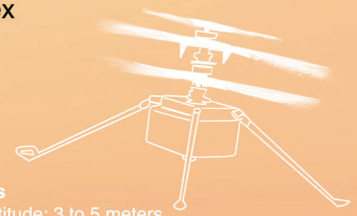
For relevant Web links, consult this issue of *American Scientist Online:*

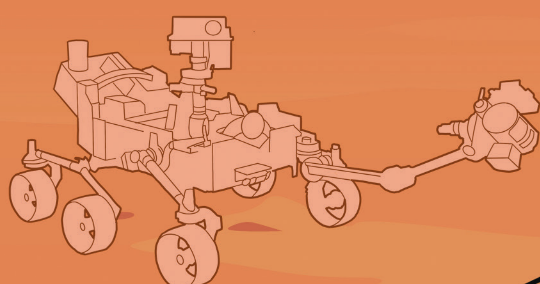
http://www.americanscientist.org/issues/id.82/past.aspx

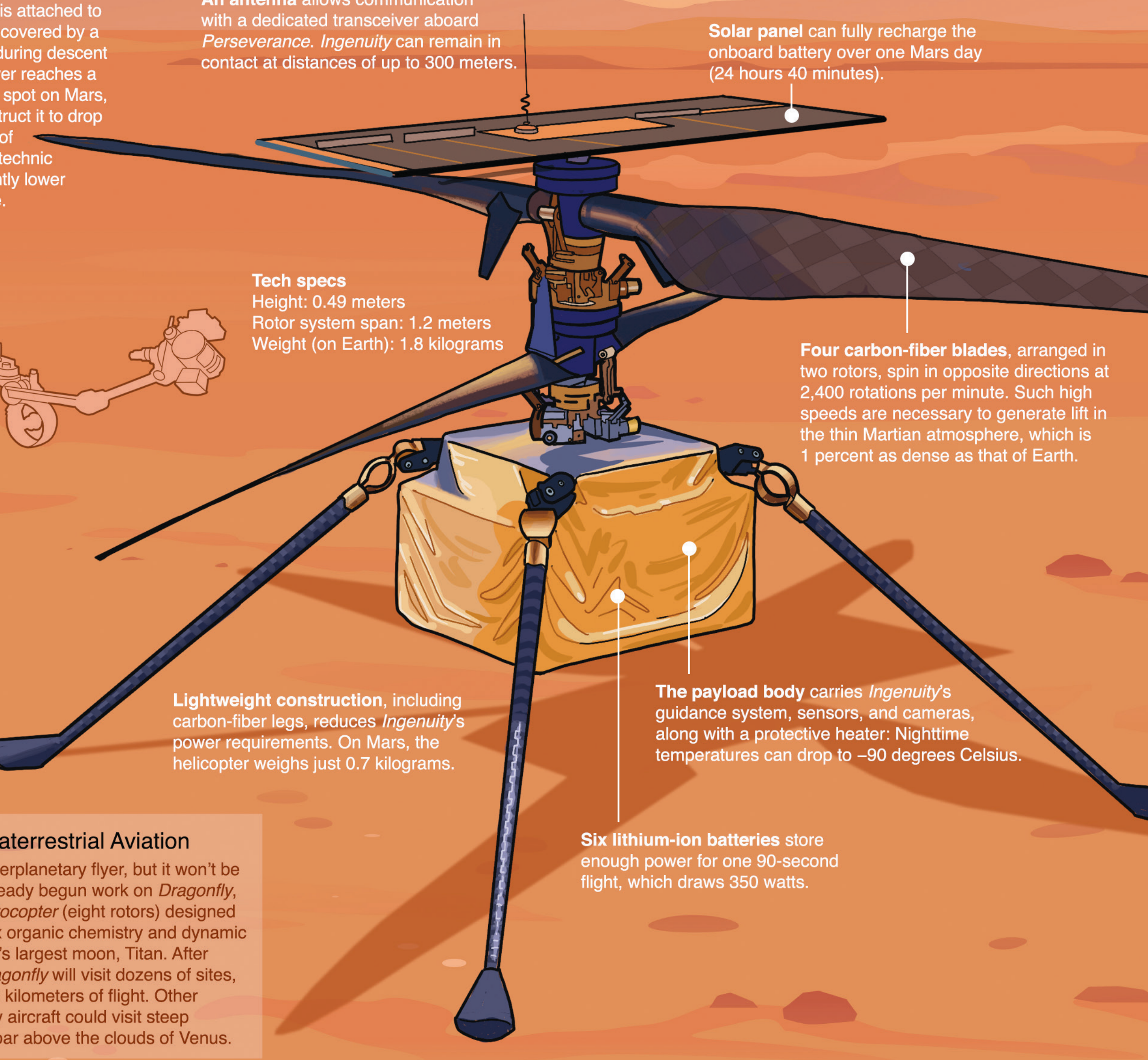
NASA's *Ingenuity* Mars Helicopter

The first attempt at powered flight on another world

A delicate helicopter named *Ingenuity* is currently headed to Mars, sharing a ride with NASA's $2.5 billion *Perseverance* rover. A few weeks after touchdown (set for February 18, 2021), *Ingenuity* will attempt the first-ever powered flight on another world. Initially, it will rise just a meter or two off the ground, hover for 20 to 30 seconds, and land. Controllers will then attempt at least four more, incrementally farther and higher flights. NASA engineers regard these forays as akin to the Wright brothers' first tentative test runs at Kitty Hawk. If *Ingenuity* succeeds, future Mars missions could include larger, more complex helicopters to scout difficult terrain or to coordinate with human explorers.







A Pioneer of Extraterrestrial Aviation

Ingenuity is the first interplanetary flyer, but it won't be the last. NASA has already begun work on *Dragonfly*, a nuclear-powered *octocopter* (eight rotors) designed to explore the complex organic chemistry and dynamic environment of Saturn's largest moon, Titan. After launching in 2026, *Dragonfly* will visit dozens of sites, logging more than 175 kilometers of flight. Other potential interplanetary aircraft could visit steep canyons on Mars or soar above the clouds of Venus.

First Person: Dante Lauretta

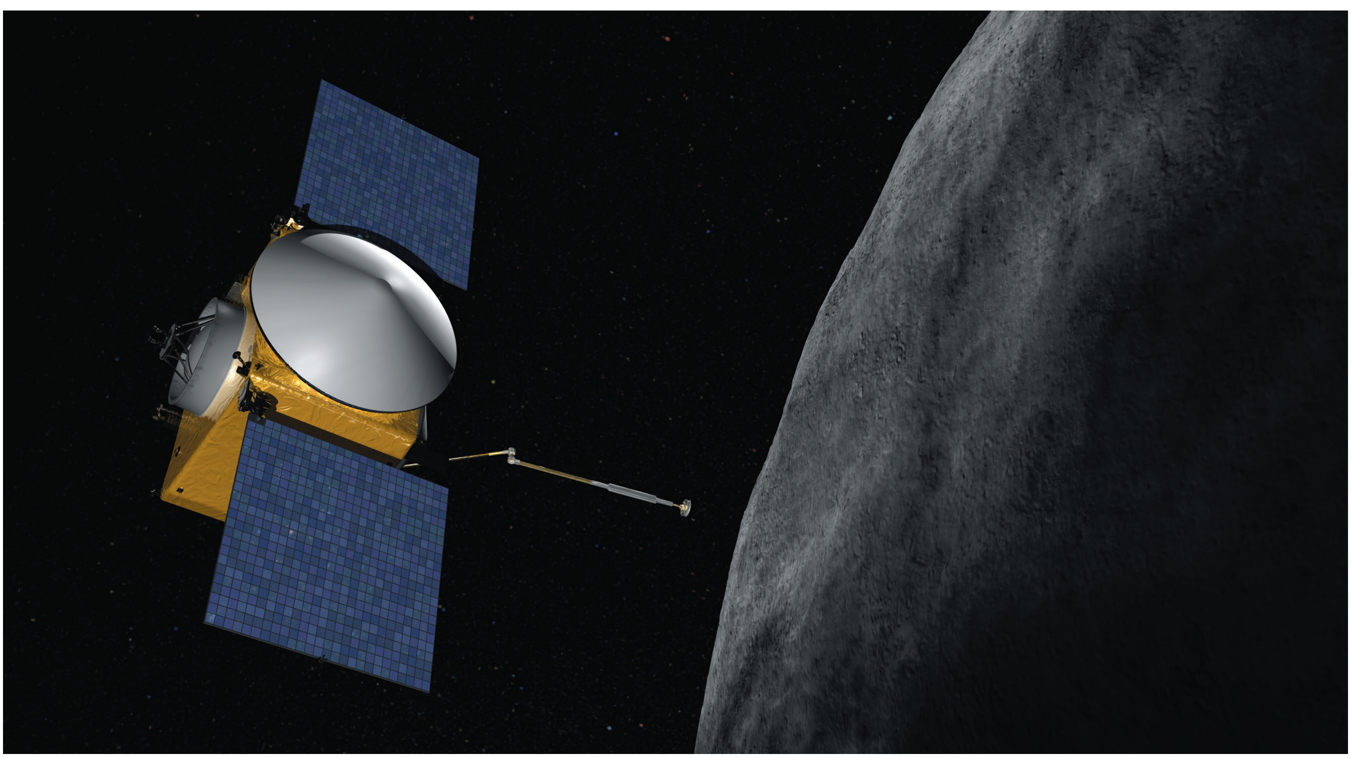
In September 2016, NASA will launch the OSIRIS-REx probe and embark on what may be the most delicate space mission ever attempted: Maneuvering next to a 500-meter-wide asteroid, brushing a robotic arm against the surface, collecting at least 60 grams of material, and flying the sample back home to a soft landing in Utah in 2023. Dante Lauretta, a planetary scientist at the University of Arizona and the mission's principal investigator, is keenly aware of both the challenges and the potential payoff. This would be the closest-ever study of an Earth-threatening asteroid, and the first chance to study up close the type of object that may have seeded our planet with the chemicals of life 4.5 billion years ago. Corey S. Powell, interim editor of American Scientist, *spoke with Lauretta about what he hopes to learn.*

Out of all the asteroids in the solar system, how did you decide where to go?
When we were making a target selection, there were around the order of half a million known asteroids. The first thing we did was say, "Well, we're not going to get out to the main asteroid belt, get a sample, and bring it back." That takes too much energy. We wanted to be solar powered, so we didn't want to get too far from the Sun. And we didn't want an asteroid that's too small or spinning too rapidly. As we looked at the data, I learned about a fascinating trend: There's a correlation between the rotation rate of an asteroid and its diameter, and there's a natural break at 200 meters. Asteroids that are 200 meters and smaller are rotating really rapidly, some of them with rotation periods of less than a minute. It looks like 200 meters is the fundamental building-block size. Most asteroids are probably rubble piles, and if you took a rubble-pile asteroid a kilometer wide and spun it that fast, it would quickly fragment into those smaller pieces.

What is so scientifically interesting about this particular asteroid?
We want to understand the organic molecular evolution of the early solar system. We want to investigate whether asteroids seeded the early Earth with the fundamental molecules that led to DNA and proteins. We're also very much interested in the origin of volatiles [compounds that vaporize at low temperatures], in particular water, and why we have so much water on Earth. If you want to understand volatile and organic evolution in the early solar system, the best place is to go is a carbonaceous primitive asteroid.

The one we selected—a 500-meter-diameter asteroid then known as 1999 RQ36, now called Bennu—also had a phenomenal data set already in place. It had made a series of close approaches to the Earth in 1999 and 2005. The Arecibo and Goldstone radio telescopes got great radar data on this object: rotation period, pole orientation, overall structure. It has one of the best known orbits

OSIRIS-REx will tap the surface and blow puffs of gas to collect samples of asteroid Bennu. (Image courtesy of NASA.)



of any asteroid in the solar system. And it turned out to be one of the most potentially hazardous asteroids, too. It has a 1 in about a 2,400 chance of impacting the Earth late in the 22nd century.

The most challenging part of the OSIRIS-REx mission will be nestling up to the asteroid and collecting a sample. Are you worried about whether the technology will work?

We recently did our first dress rehearsal and stepped through the whole sample acquisition sequence—it's about five hours from when we decide we're going for the sample to when we contact the asteroid surface. It started to dawn on me how nerve-wracking those five hours are going to be. We're hoping to get it in one shot, but we can get three shots if needed.

This will be the first time a spacecraft has brought a sample of an asteroid back to Earth. What will a direct sample tell you that you can't learn remotely?

OSIRIS-REx will allow us to map the distribution of water and organics in the inner solar system. That will be important for people who are thinking about going out to these asteroids as resources for use in human exploration. It will also allow us to get a much better understanding of the distribution of the kind of material that's preserved there, and tie that back to our models for the formation of the solar system. We're trying to compare meteorite spectra to asteroid spectra to understand why they're different—and they really are different.

Studying the formation and delivery of organic material to the Earth is really hard with any meteorite that fell on this planet because of contamination. It's why a sample return mission is so critical.

OSIRIS-REx will also explore the structure and dynamics of asteroid Bennu. What are your goals there?

One of the things we're excited about is that we're going to get up close with a rubble-pile asteroid and really get into the surface geology. If a rubble-pile asteroid comes close to the Earth it gets stretched out, almost like a cigar, and then it will snap back. That may be why Bennu has a spinning-top shape. That process may also generate particles as a result of friction, and all that material should be accumulating at the equator. I've challenged my science team to figure out where this asteroid originally came from. How did it get into the inner solar system? How many planetary close encounters has it had during its lifetime? Then we're going to get a piece of it back on Earth and run all kinds of tests that are going to tell me the history of this material. How long has the surface been exposed to space? When was the last major impact on this asteroid? We're going to test—for the first time—the geologic and dynamic theories of asteroid evolution.

How do you study the geologic history of an asteroid?

We're going to track it very precisely for six days, and we'll get a very nice map of the gravity field. We're going to image the asteroid from all kinds of illumination angles and build up a three-dimensional model based on how the shadows change. From the shape and gravity field, you can tell if it's homogeneous or if there's a density variation inside. We'll also look at any surface expressions of internal structure. Do we see faulting, ridges, scarps, anything that would give us some insight into deeper geophysical occurrences?

Your current thinking is that Bennu is the debris from a collision between two larger parent bodies, is that right?

Yes. We're looking at three asteroid families in the inner main belt as the most likely source. These families are a result of major collision between two large bodies in the main asteroid belt [between Jupiter and Mars], anywhere from 200 million to 2 billion years ago. The collisions take two asteroids and shatter them to thousands and thousands of fragments.

One of the big surprises to me is that sunlight can move asteroids by heating the surface: Thermal radiation exerts a small push that can change shift its orbit (called the Yarkovsky effect). How does that affect Bennu?

What happens is the smaller you get, the more the Yarkovsky effect changes the semi-major axis [the size of the orbit]. Yarkovsky appears to be a size sorting mechanism in the main asteroid belt, where the smallest asteroid from a collision like the one that formed Bennu gets pushed really quickly and delivered into the inner solar system. This size sorting effect explains the size distribution of asteroids in the inner solar system.

What if the day comes when we find an asteroid that has a high likelihood of hitting Earth? Will OSIRIS-REx help figure out how to avoid an impact?

Absolutely. We're building a spacecraft that's going to launch from Earth, rendezvous with an asteroid, characterize its fundamental properties, and ultimately descend to the surface in a series of precision maneuvers to a spot of our choosing. Any kind of deflection where you want to rendezvous with the asteroids is going to require those techniques. Ultra-fine thrusting in microgravity—it's never been done before. That's the first critical thing.

The second thing is we're going to measure directly the Yarkovsky effect, which is the largest uncertainty in orbit propagation into the future. If you've got a couple of decades before an impact is going to occur, you can actually use the Yarkovsky effect—you can direct it. You could paint some areas of the asteroid white, some areas black. You could control that Yarkovsky force, but only if the theory matches the observed acceleration. We're going to test that. I think that is one of our most valuable contributions to an impact hazard mitigation.

Inside Jupiter

NASA's Juno spacecraft is mapping the history of our Solar System in our giant neighbor.

Scott Bolton

Juno is a true mission of exploration and discovery. The data from the spacecraft have been paradigm-shifting for our understanding of giant planets, providing a revolutionary new view of Jupiter, both challenging our theories and presenting beauty that is almost beyond belief. The images returned from Juno are breathtaking, bridging art and science.

Although Juno represents many firsts for NASA, one of the most interesting is how the spacecraft's camera, JunoCam, is set up to involve the public in the mission. The Juno team has chosen to make all of Juno's raw, unprocessed imaging data available to the public via the mission's website. This decision not only allows, but rather requires, that all of the pictures of Jupiter taken by Juno are essentially created by the public, because there is no official JunoCam science team. The citizen scientists involved are not modifying NASA images, they are creating the images themselves. They are literally the first humans to see Juno's discoveries. Jupiter's giant polar cyclones, the first high-resolution close-up of the shrinking Great Red Spot, the high-altitude pop-up clouds discovered near the edges of the planet's swirling storms—these are just a few of the discoveries made because of imagery created by citizen scientists. And citizen artists are also involved, creating artwork inspired by Juno's exploration of Jupiter.

Jupiter is both a record and a driver of the formation of the planets. It provides a glimpse into the earliest stages of the development of our Solar System.

Having launched on August 5, 2011, from NASA's Kennedy Space Center, Juno arrived at Jupiter on July 4, 2016. The journey had started 13 years earlier, when a team of scientists from around the world began meeting to discuss what is arguably one of the most challenging and scientifically ambitious planetary missions NASA has ever attempted. Juno was designed to answer questions raised by knowledge gained over the history of space exploration. Previous missions demonstrated the diversity of the planets and brought up fundamental questions regarding the vulnerability of Earth, how life started, and the origin of our Solar System—topics that rose to the top of NASA's priorities. Juno was intended to fill a huge gap in our understanding of planetary formation. Previous exploration of Jupiter had shown us just how important this celestial body was, but it also left us with unanswered puzzles

QUICK TAKE

The Juno spacecraft was built to handle the heavy radiation around Jupiter: It has a central vault to protect its equipment and a flight path that minimizes exposure.

Many theories about Jupiter have not been borne out by the data from Juno. Its data indicate a larger, more diffuse core, and a more heterogeneous atmosphere than expected.

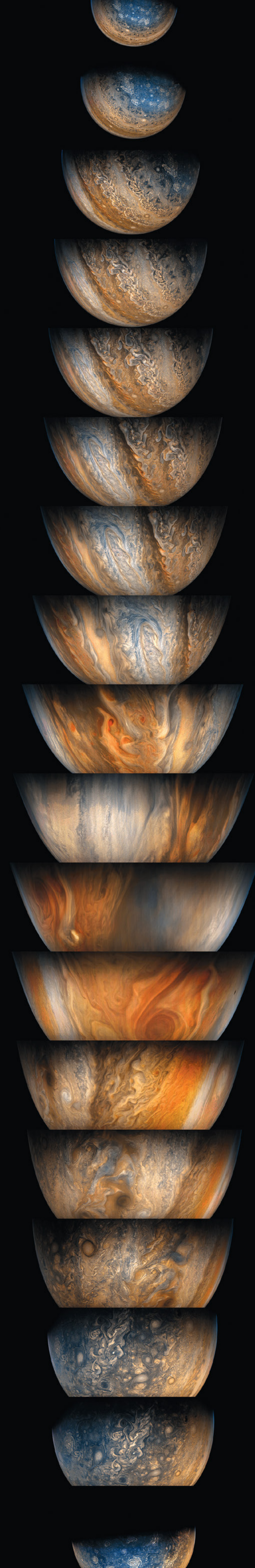
Jupiter was the first planet to form in the early Solar System, and it used up more than half of the available matter; understanding its composition can help the study of exoplanets.

NASA/JPL-Caltech/SwRI/MSSS/Kevin M. Gill

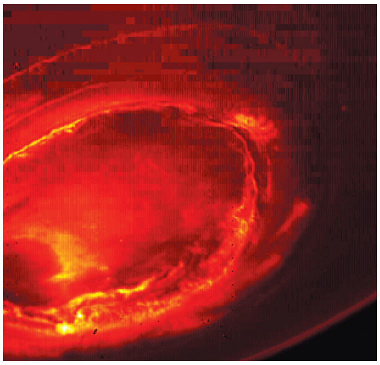
related to how Jupiter formed and how its composition was different from that of the Sun. The history and distribution across the early Solar System of water and other volatile compounds vital to Earth and to life were cloaked in mystery. Scientists knew that Jupiter was no ordinary planet, and that it might hold the keys to understanding the creation of the rest of the Solar System.

Jupiter is larger than all of the other planets combined, and its formation shaped the composition of the rest of the Solar System. The planet's gigantic mass literally threw material all around the early Solar System, including the key ingredients of life—water and organics—but also rocky material. In this way, Jupiter determined not only when Earth formed, but also what the planet is made of: Jupiter's presence is what made Earth habitable. Learning how Jupiter formed provides a unique glimpse into the earliest stages of our Solar System. Jupiter is also our archetype for extrasolar giant planets, so what Juno learns not only helps us understand our own origin, but also illuminates how planetary systems are formed around other stars as well. Truly, the history of our Solar System is recorded in the formation of Jupiter.

THE JUNO SPACECRAFT IS named after the Roman mythological goddess who saw through clouds obscuring her husband's mischief, because the spacecraft's instruments were designed to peer past Jupiter's cloud layers and investigate its internal workings. The visible-wavelength images have revealed swirling patterns and storms (*such as those above*), but the craft's eight scientific instruments have observed Jupiter and its associated phenomena in many other ways. Juno cartwheels as it orbits Jupiter to increase its stability, and it takes an extreme polar path, skimming just above the cloud tops, to avoid the most intense areas of radiation. Its orbits have created pole-to-pole time-lapse views of Jupiter, such as the one at right: The north pole is at top; the south pole at bottom was imaged two hours later. The images in between capture the equatorial bands of belts and zones.



NASA/JPL-Caltech/SwRI/MSSS/Gerald Eichstädt/Seán Doran

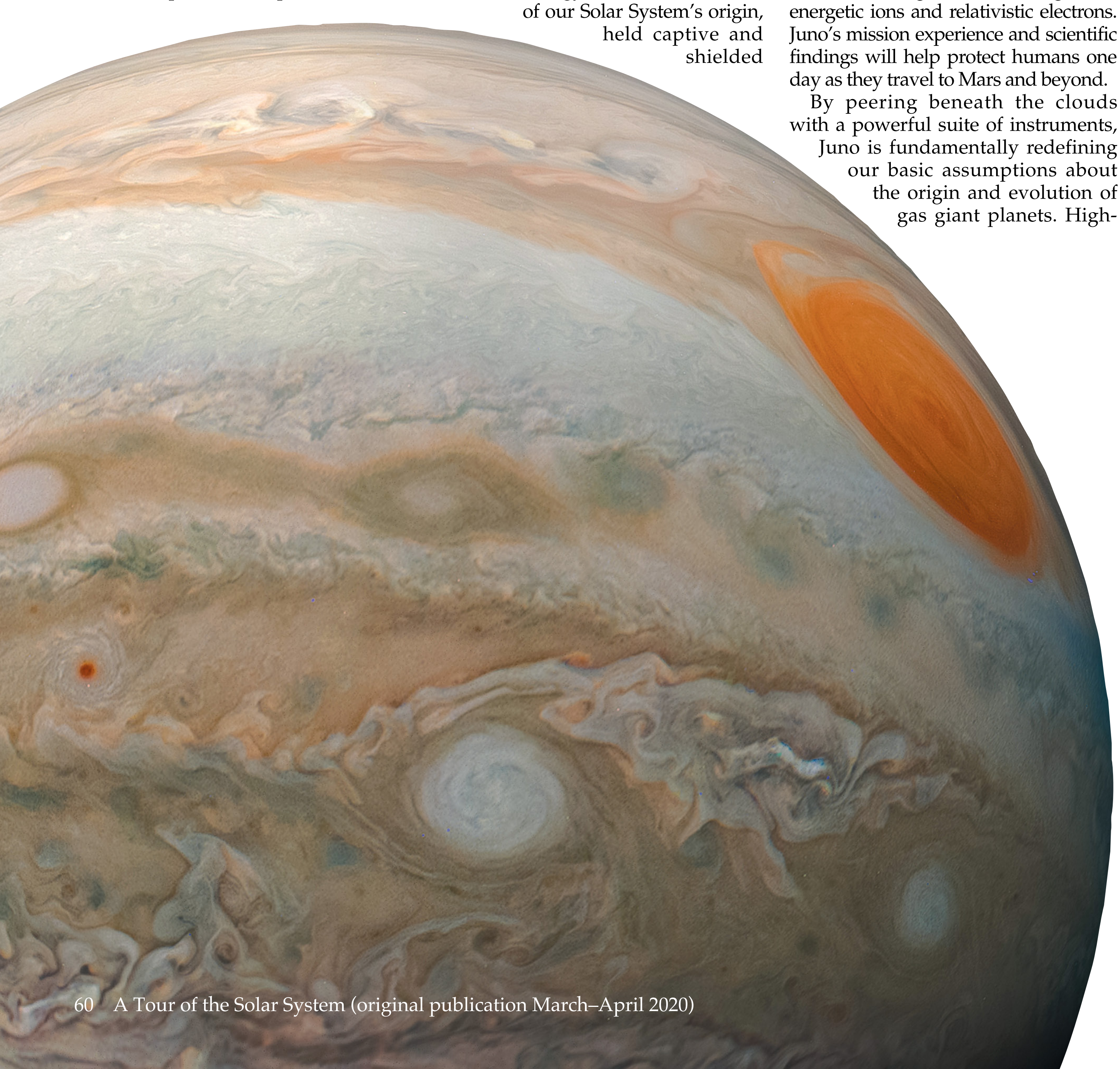


NASA/JPL-Caltech/SwRI/ASI/INAF/JIRAM

THE SOUTHERN POLE of Jupiter experiences intense auroras that are largely unseen from Earth, because of the angular misalignment between the two planets. But Juno's polar orbit allows it a perfect view of these events, captured here in infrared. Jupiter's auroras are the most powerful in the Solar System, and initial results from Juno indicate that the mechanism producing them is more complicated than that on Earth.

By design, Juno would enter into one of the harshest environments in our Solar System. A magnetic swarm of highly energized charged particles surrounds Jupiter. These particles are so fierce that they are moving at nearly the speed of light: They are capable of penetrating our strongest shielding and can destroy even our most advanced electronic technology. To reveal the secrets of our Solar System's origin, held captive and shielded from view by Jupiter, a vehicle had to be created that was more armored tank than spacecraft. The scientists and engineers working on this goal blended some of the most advanced shielding technology with new types of instruments capable of seeing the invisible layers below Jupiter's stormy clouds, and they carefully designed a flight path to thread the needle, finding a route that avoided the most intense radiation and penetrating particles. Juno is the first spacecraft with a radiation shielding vault, pioneering a key tool that will also be needed for future longer-duration human space missions. Radiation protection is the primary challenge to sustained human exploration beyond Earth. Jupiter is a major source of this radiation, filling its surroundings with energetic ions and relativistic electrons. Juno's mission experience and scientific findings will help protect humans one day as they travel to Mars and beyond.

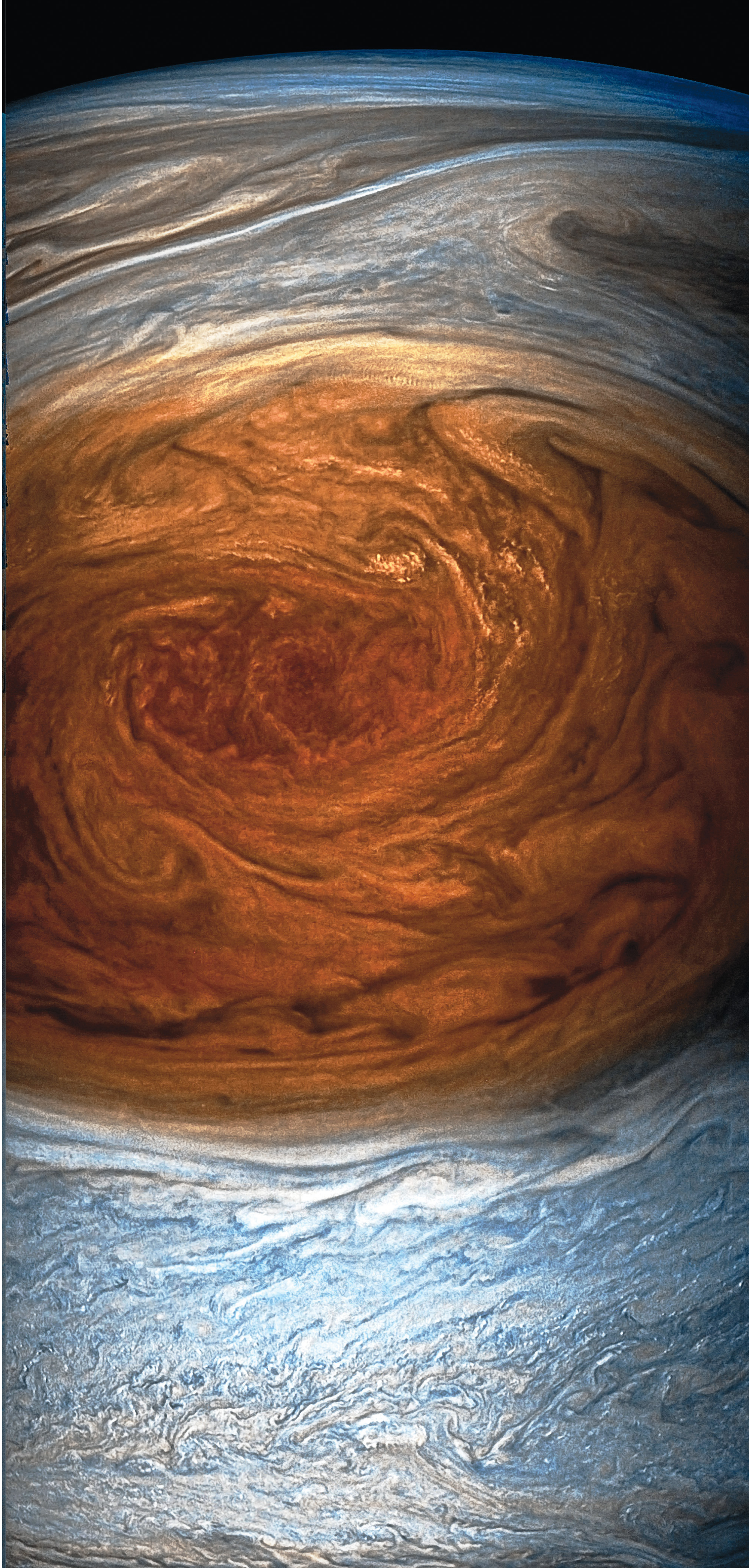
By peering beneath the clouds with a powerful suite of instruments, Juno is fundamentally redefining our basic assumptions about the origin and evolution of gas giant planets. High-



NASA/JPL-Caltech/SwRI/MSSS/Kevin M. Gill resolution imagery returned by Juno's camera has revealed myriad Earth-size cyclones raging in Jupiter's atmosphere. Microwave measurements have discovered layers of ammonia clouds stretching to great depths. The atmosphere is not homogeneous, which fundamentally challenges our ideas of how giant planet atmospheres work. Juno sees deep within Jupiter into a metallic hydrogen region, searching for evidence of a compact core. Surprisingly, Juno discovered that Jupiter's core is fuzzy, without sharp boundaries, opening up new theories to explain this giant planet's formation and evolution. Possibly, Jupiter suffered a large impact early on, similar to what scientists have theorized about how our Moon was created from an early impact that Earth experienced. Jupiter's incredibly strong magnetic field had its own surprise: a giant magnetic anomaly called the Great Blue Spot, near the equator.

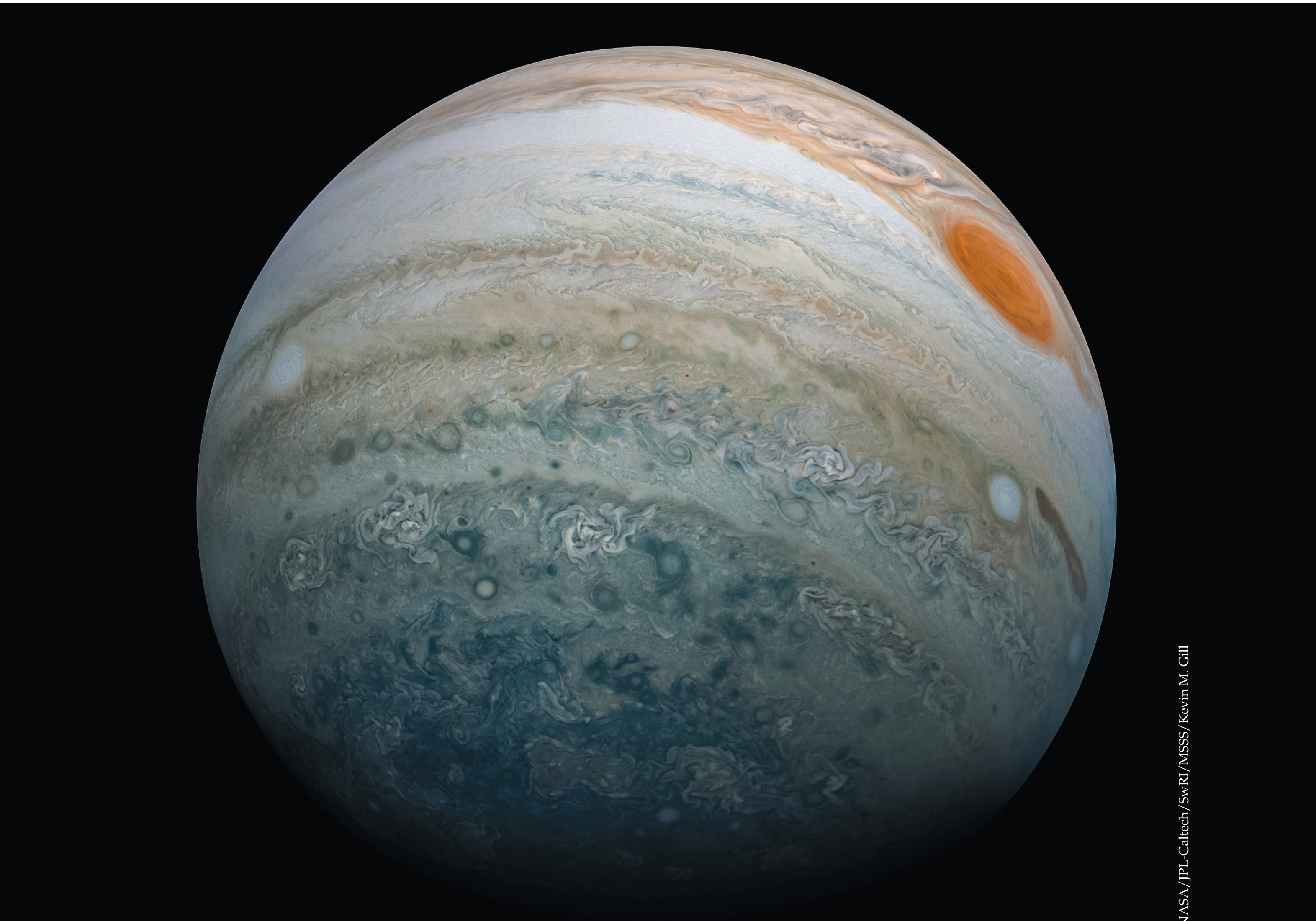
The images returned from Juno are works of art and wonder. But the science results from the mission are simply dramatic.

THE GREAT RED SPOT on Jupiter (*left, and close up, right*) has been a prime target for Juno imaging. Scientists have wanted to know how deep are the roots of Jupiter's most famous landmark. Juno data indicate that this storm, about 1.3 Earths wide, has winds that penetrate at least 300 kilometers into the planet's atmosphere. The spot has been shrinking since it was first drawn and photographed more than a century ago. In 1979, when NASA's Voyagers 1 and 2 passed Jupiter, it was roughly twice Earth's diameter. Juno provides a close-up view of the dynamics that are occurring as this centuries-old iconic feature of Jupiter keeps changing. Juno has seen "streamers" protruding from the great storm, as seen at left, which flake off and dissipate, and may be related to why the spot is shrinking in latitude.



NASA/JPL-Caltech/SwRI/MSSS/Gerald Eichstädt/Seán Doran

NASA/JPL-Caltech/SwRI/MSSS/Björn Jónsson

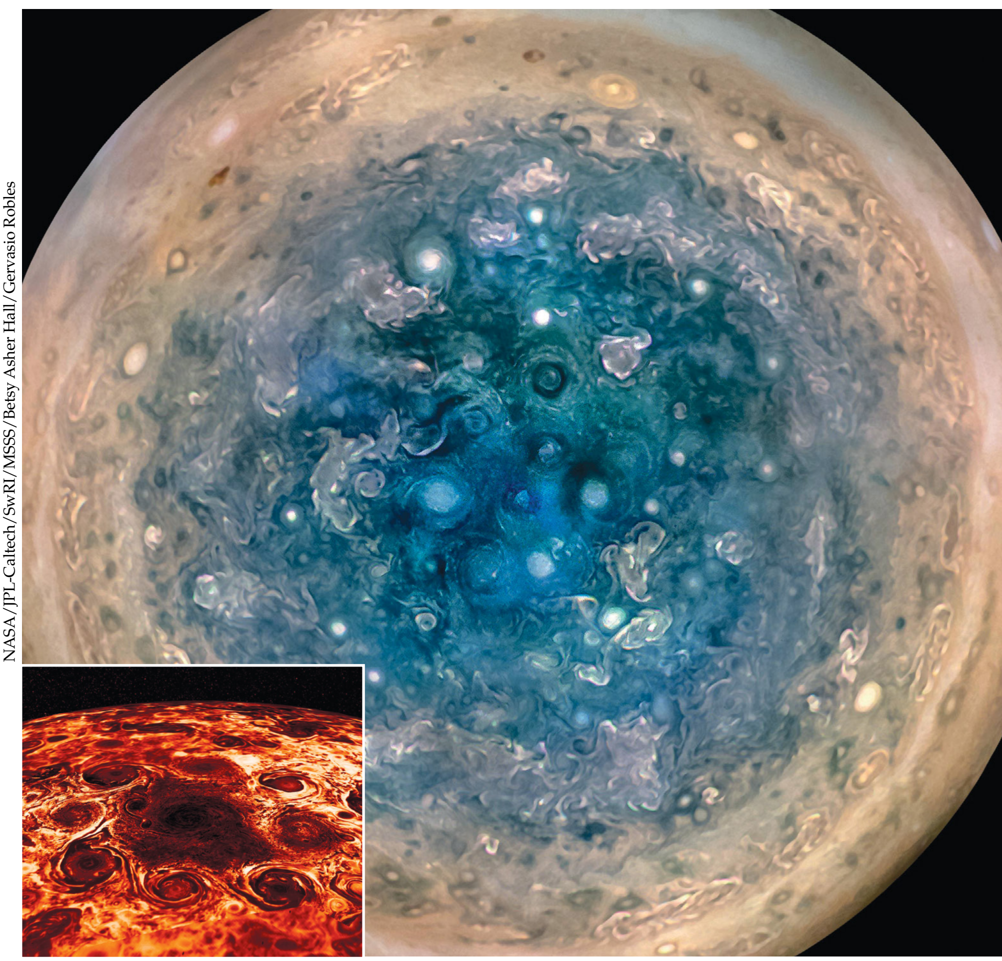


NASA/JPL-Caltech/SwRI/MSSS/Kevin M. Gill

NASA/JPL-Caltech/SwRI/MSSS/Gerald Eichstädt/SeánDoran

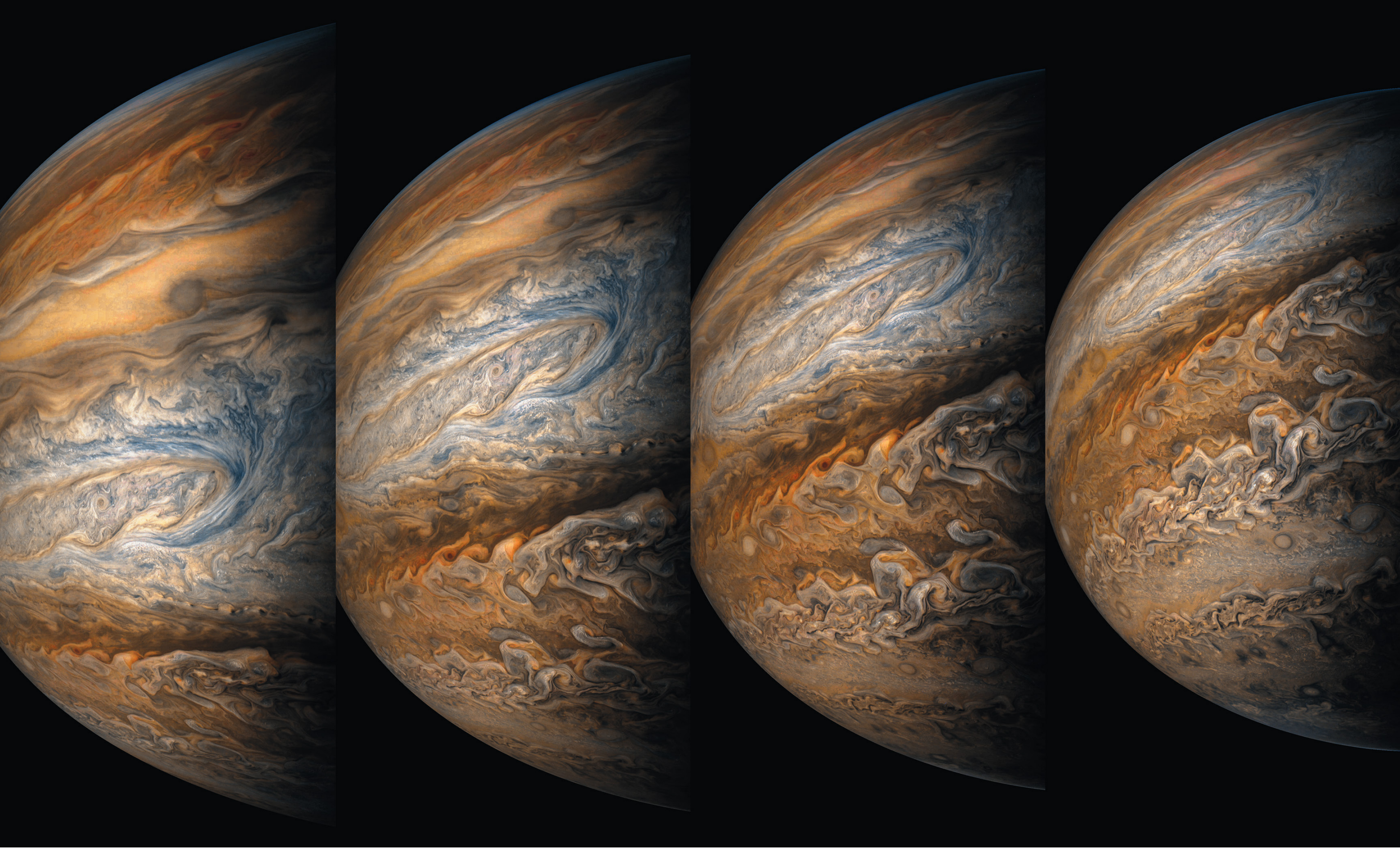
ASSESSING THE AMOUNT and distribution of water in Jupiter's clouds (*left*) and interior is another goal of the Juno mission. Water and water ice existed in the nebula from which the planets were formed, so its study can help in understanding how close to the Sun Jupiter was when it formed, and the process by which Jupiter's composition came to differ from that of the Sun. Juno has found that Jupiter's water and ammonia are variable across the planet and at great depth, a surprise to scientists that complicates theories on how atmospheres work on all giant planets. Results from Juno indicate that water is enriched, compared with water in the Sun, at Jupiter's equator. Over the next few years, Juno's path will get closer to Jupiter's northern latitudes, allowing scientists to be able to compare this result with the rest of the planet, and learn about both Jupiter's deep atmospheric meteorology and its formation.

IMAGES DO NOT ROLL OUT of Juno's camera fully formed. The spacecraft captures narrow strips with three color filters, which have to be combined. Most of the dramatic images from Juno were created by citizen scientists. Although Jupiter has many colors (*examples at left*), most images are color-stretched to improve clarity and identify atmospheric details that would not otherwise be easily apparent (*above*). Scientists believe that darker-colored clouds lie deeper in the atmosphere. Roiling storms (*white*) are thought to be driven by warmer gases welling up from the interior to the surface; that temperature gradient, combined with the planet's rotation, creates atmospheric circulation and wind speeds of several hundreds of kilometers an hour. Juno has also discovered that both poles are surrounded by clusters of cyclone-like storms, five around a central one at the south pole (*right*), and eight around one at the north pole (*inset, in infrared*).



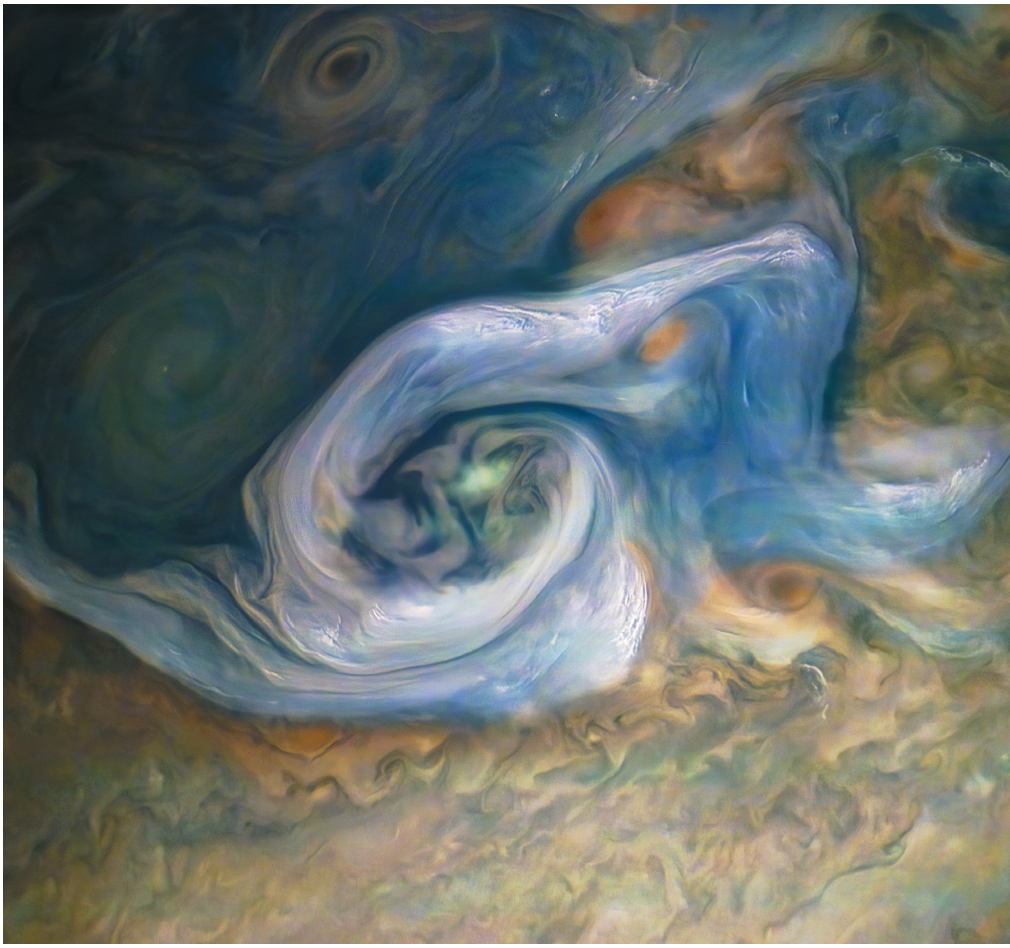
NASA/JPL-Caltech/SwRI/MSSS/Betsy Asher Hall/Gervasio Robles

NASA/JPL-Caltech/SwRI/ASI/INAF/JIRAM

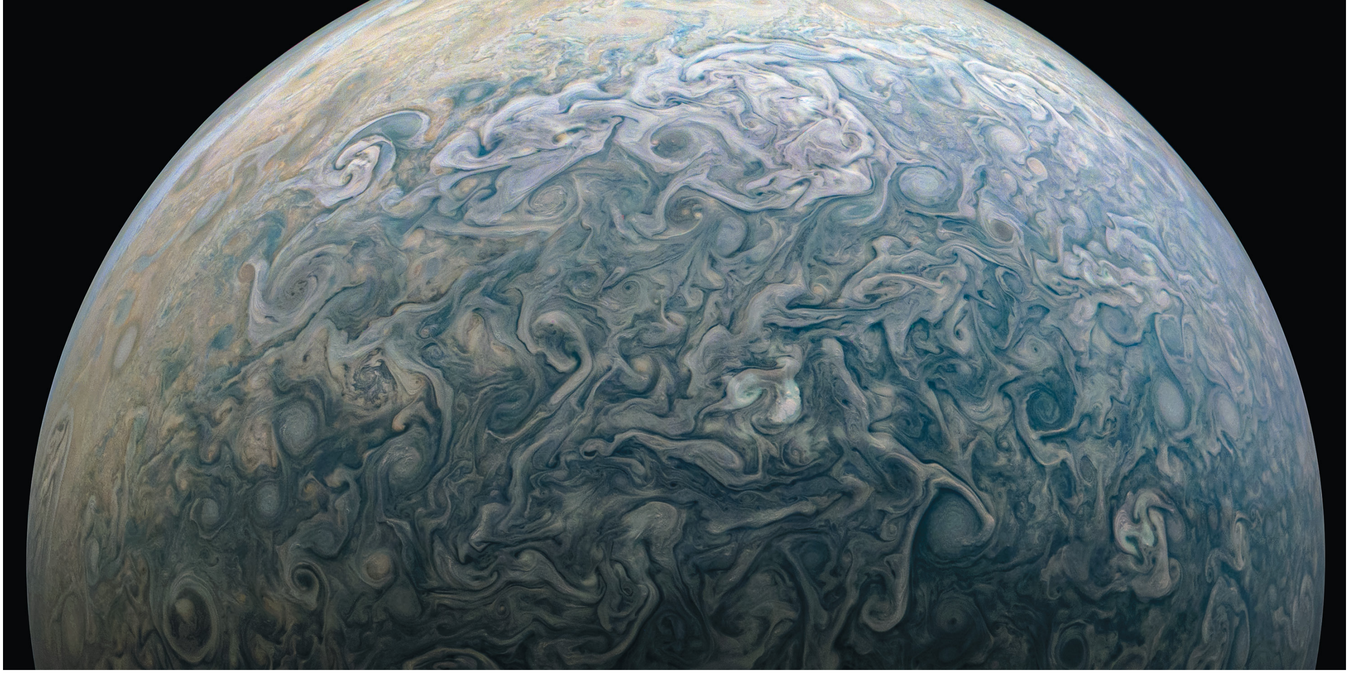


NASA/JPL-Caltech/SwRI/MSSS/Gerald Eichstädt/SeánDoran

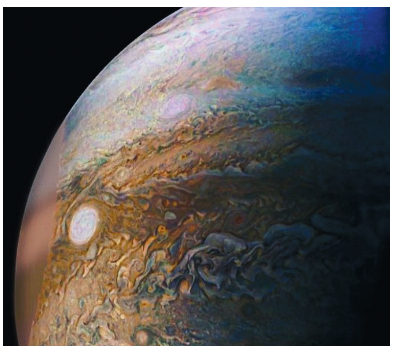
JUNO TAKES 53 DAYS to complete an orbit of Jupiter, but it obtains most of its images (*such as those above*) during an eight-hour window when it is closest to the planet, as near as 3,400 kilometers above the cloud tops. Jupiter's atmosphere is thought to have at least three cloud layers: The top is likely composed of ammonia ice, the middle of ammonium hydrosulfide crystals, and the bottom likely of water ice and vapor. Juno probes for trace gases, especially ammonia and water, to study atmospheric circulation. Previously it was thought that ammonia distribution was uniform down to deep levels, with variability only in the uppermost regions, where ammonia clouds form (*like the ones in the storm at right*). Juno discovered that Jupiter has an ammonia-rich band concentrated in a narrow belt around the equator and is greatly depleted in other regions, showing a surprising depth to its atmospheric circulation.



NASA/JPL-Caltech/SwRI/MSSS/Gerald Eichstädt/SeánDoran



NASA/JPL-Caltech/SwRI/MSSS/Kevin M. Gill



NASA/Juno

ALTHOUGH THE SWIRLING CLOUDS on Jupiter are dramatic, some of Juno's most surprising findings are not visible. Jupiter's core cannot be examined firsthand, even by a probe, because the pressures inside the planet are so intense. So Juno studies Jupiter's core indirectly by measuring the planet's magnetic and gravitational fields. Because of the planet's fast spin (a Jupiter day lasts 10 Earth hours), it bulges at its equator, which alters the gravitational pull on Juno as it orbits the planet, affecting the spacecraft's speed. Measurements of this speed shift have been used to create a map of Jupiter's gravity, which has told researchers that the relatively massive core also has a large volume, surprisingly, meaning it must be larger and more diffuse than previously thought, rather than being compact. The Juno mission will continue until at least mid-2021, providing opportunities to discover more surprises that Jupiter has hidden away.

JunoCam/NASA/Synthetrix

Scott Bolton is the principal investigator for the Juno mission and associate vice president of the Space Sciences and Engineering Division at the Southwest Research Institute. He earned his PhD in astrophysics from the University of California at Berkeley in 1990. He led the concept design and the development of the Juno spacecraft, as well as the creation of its microwave radiometer experiment. His research expertise includes microwave radio astronomy, atmospheric science, space physics, and the origin of the Solar System. He previously served as chair of the Cassini Orbiter Titan Science Team, and of the Magnetospheric Science Working Group for the Galileo mission, in addition to other missions. Email: scott.bolton@swri.org

For relevant Web links, consult this issue of *American Scientist Online:*

www.amsci.org/magazine/issues/2020/march-april

Tides and the Biosphere of Europa

A liquid-water ocean beneath a thin crust of ice may offer several habitats for the evolution of life on one of Jupiter's moons

Richard Greenberg

The Copernican revolution began over 500 years ago with the realization that the Earth was not the center of the universe, but we still await its grand finale: The anticipated discovery of life elsewhere. Where else might we find life? The vast scale of the universe makes it virtually certain that there are other Earth-like settings. In our own solar system, Mars's distance from the Sun makes it sufficiently Earth-like that, especially with increasing evidence for occasional liquid water, many are looking there for the first signs of extraterrestrial life. Recently, however, a new contender has emerged, and surprisingly it is from the cold outer solar system: It is Jupiter's moon Europa.

Europa played an early role in the Copernican revolution as well. As one of the four satellites of Jupiter discovered by Galileo in 1610, Europa provided evidence that objects could orbit a celestial body other than Earth. Its steady motion around the planet, observable in the smallest backyard telescope, has been followed ever since. However, it wasn't until 20 years ago that scientists realized that the tidal stresses imposed on Europa by the giant Jupiter could generate enough heat to maintain water in a liquid state, even so far from the Sun.

Richard Greenberg is a professor of planetary sciences at the University of Arizona's Lunar and Planetary Laboratory. His research involves the application of celestial mechanics to understanding the physical processes, characteristics and history of the planets. As a member of the Galileo imaging team, and with his interdisciplinary research group at the University of Arizona, he has applied his expertise to understanding how tidal processes can explain what we see on Europa and what habitable settings may lie below the surface. Address: Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ 85721. Internet: greenberg@lpl.arizona.edu

The possibility of liquid water opened the door to speculation about life, but water alone is not sufficient to sustain organisms. Life requires a nurturing environment, appropriate chemistry, a source of energy—a habitat. Now, late-20th-century observations by another Galileo—this time the Galileo spacecraft in orbit around Jupiter—show that tidal processes may create physical conditions that support a variety of interconnected habitable settings.

The extent to which tides govern the physical nature and potential habitability of Europa was quite unexpected. I have been part of the Galileo imaging team since the project began in 1977, and although I suspected that the Jovian tides would play a role in the rotation of the satellites, and perhaps govern some of their geophysical processes, no one fully anticipated the role that tidal processes would play in everything we see on these moons. Nor did I expect to find myself in the astrobiology business—which, for the moment, doesn't study extraterrestrial life per se, but rather concerns itself with where alien life might be found. Here I recount some of what we've learned about Europa, and how these observations hint at the existence of habitable environments.

Waterworld

Planetary scientists have known for decades that Europa's surface is predominantly water ice, beginning with ground-based spectroscopic studies by Gerard Kuiper, among others. The density inferred from gravity measurements suggests that the water layer may extend down as far as 150 kilometers below the surface, or somewhat less if part of the low-density layer is clay under the ocean. Although the surface is frozen, below it most of the water layer is probably liquid.

To the human eye, Europa would appear white and bland. Most images are processed with the color and contrast exaggerated to reveal surface features. In that way, even a distant view (Figure 1) manifests the two main types of geological terrain: The lines represent tectonic features, such as cracks and rifts and ridges, and the splotches represent disrupted, chaotic terrain. In fact, nearly all of Europa is covered by either tectonic or chaotic terrain, in roughly equal measure. Both appear to have been formed by processes driven by tides: Over the 85-hour day on Europa, tidal distortion creates stress that correlates well with tectonic features. And the chaos likely results from modest local and regional concentration of the tremendous internal heat of tidal friction.

The cracks and ridges and chaos make for a surface far too rough for a hockey game, and too challenging for any ice rink's Zamboni. But neither can the weak ice support high mountains. Topography is at most a few hundred meters high.

Perhaps the most significant feature in any full-disk view of Europa is what we do not see: craters. In fact there are a few, such as Pwyll, the crater whose bright rays of splashed ice extend for 1,000 kilometers in every direction (Figure 1). But contrast the scarcity of craters with the heavily bombarded surface of our Moon, or with Europa's neighbor Callisto. Given the large number of tiny bodies in the outer solar system, especially comets, Europa's surface must be young to have avoided heavy bombardment. The tidal processes that drive tectonics and chaos have been so active recently that they have

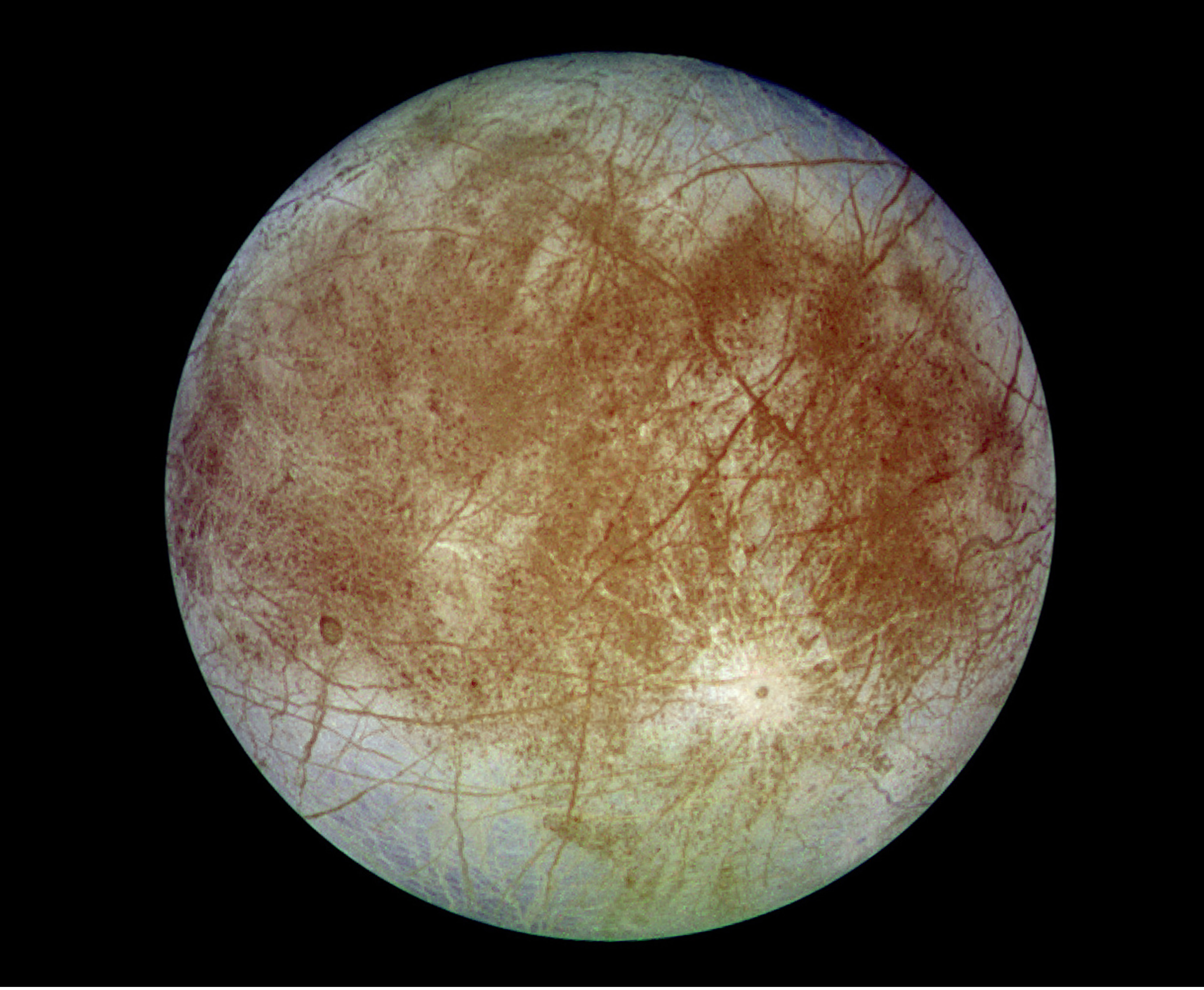


Figure 1. Europa's icy crust is marked by two types of geological terrain: tectonic, which appears here as lines, and chaotic, which appears as splotches. Both types of features appear to have been formed by processes driven by Jupiter's tidal forces. Recent interpretations of these features suggest that the icy crust may be relatively thin, and that a liquid-water ocean may lie close to the surface. If so, life on Europa may be able to exploit several habitable environments. Europa is about 3,100 kilometers in diameter, roughly the same size as Earth's moon. The color and contrast in this image are exaggerated to reveal surface features. (Except where noted, all images were made by the Galileo spacecraft, and are courtesy of the Jet Propulsion Laboratory and NASA.)

completely resurfaced Europa in the cosmically short time since dinosaurs vanished from Earth. It would be surprising if they are not still in action.

How could tidal processes form the terrains on Europa's surface? I discuss next how these processes seem to involve interaction of the surface with the ocean below, producing a variety of habitable niches. Comfortable niches would be stable for thousands of years, but individual niches would come and go over longer times. Not being too secure, organisms would need to adapt to a continuously habitable but ever-changing world, an essential driver for evolutionary advancement of life.

Tides

Just as the Sun and the Moon tend to elongate the Earth along an axis directed toward them, Jupiter elongates the shape of Europa along an imaginary line connecting the satellite to the giant planet. On the Earth, most of the continual reconfiguring is done by sloshing of the oceans, although the solid body of the Earth is also worked to some extent. Similarly on Europa, the solid body undergoes distortion, but most of the amplitude of tidal change is due to change in the shape of the liquid ocean under the ice.

Tides on Europa are different from those on the Earth in important ways. For one thing, Jupiter is huge, and it produces enormous tides on Europa (Figure 2). The height of the tide is about 500 meters at its peak on both sides of the moon. However, since Europa rotates nearly synchronously, keeping the same face toward Jupiter for hundreds of years, the daily tidal change is much smaller. The length of Europa's day matches its orbital period of 85 hours, and the only reason the tide changes over the course of a Europan day is that

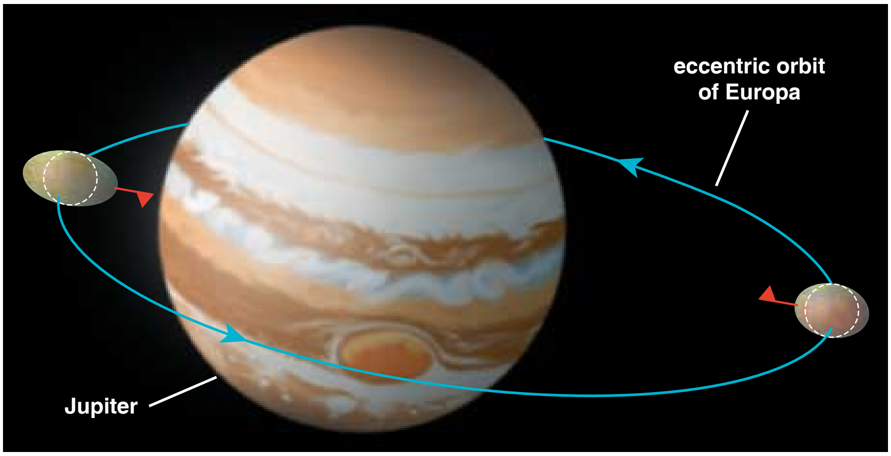


Figure 2. Europa's tidal shape changes as it moves along its eccentric orbit around Jupiter (left). The Europan tides are highest when the moon is closest to Jupiter, and lowest when it is farthest away. The same side of Europa (red flag) faces Jupiter throughout the moon's 85-hour orbit. Here the orbital shape and the relative sizes of the bodies are greatly exaggerated. In reality (right), Jupiter is considerably larger than Europa (white ball, to lower left of the Great Red Spot). Here Europa lies about 600,000 kilometers above the Jovian clouds.

its orbit is eccentric. At the time each day when Europa is closest to Jupiter the tidal distortion is greatest, raising the tides by an extra 30 meters, and when Europa is farthest from Jupiter the tide decreases by 30 meters. This daily working of the surface is what creates frictional heat and stresses the icy crust on top of the liquid ocean.

The daily tide has another important effect. Because of its distorted shape, the elongated ends of Europa are pulled by Jupiter in such a way that it drives the moon's rotation to a rate slightly faster than synchronous. As a result, the large 500-meter high tide moves around Europa over the course of tens of thousands of years. Over days or a few years, however, only the 30-meter diurnal tide works the satellite.

Tides also tend to damp down orbital eccentricities very quickly. How then does Europa's orbit remain so eccentric? The answer goes back to a remarkable feature of the orbits of the Galilean satellites, evident even in Galileo's 17th-century observations. During the 85-hour orbital period of Europa, satellite Io completes exactly two orbits and Ganymede makes exactly half an orbit. Two centuries ago, Laplace demonstrated how this 1:2:4 ratio of periods could resonantly enhance the mutual gravitational effects of the satellites, so that they keep one another's orbits eccentric and maintain the whole-number ratio of periods. The orbital resonance is critical in maintaining the tides that heat and stress Europa, as well as Io (even more) and Ganymede (considerably less so).

Cracks and Ridges

As the satellite changes shape under the influence of tides, the thin ice shell riding over the surface is stressed. The large-scale linear patterns on Europa correlate roughly with theoretical tidal tension, suggesting that the lines, or lineaments, represent cracks in the shell.

The Galileo spacecraft's camera zoomed in on selected locations, so we can see how these cracks manifest themselves at higher resolution. Figure 3a shows the intersection of two major global lineaments, just north of the splotch known as Conamara Chaos, with a resolution (pixel size) of about 1 kilometer. We find that most global lineaments at this scale prove to be double dark lines, with a bright gap between them. These features were named triple bands when they were discovered on Voyager images. Using similar terminology, the double line down the center of a highway would also be a triple band, if one counts the space between the lines.

Viewing the same region at higher resolution, with a lower illumination angle that shows topography and landforms, we find a very different picture

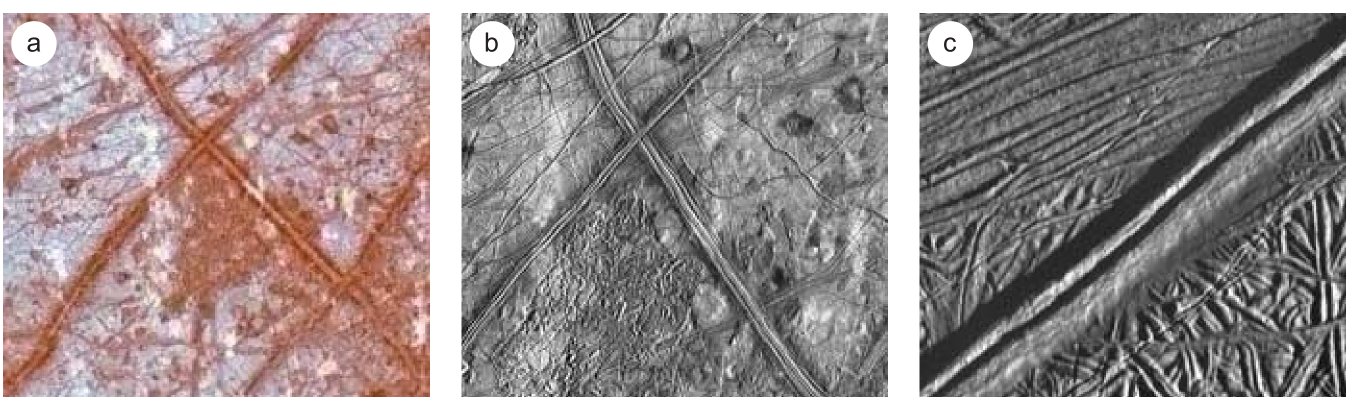


Figure 3. Triple-band features, such as the global lineaments (X-pattern) that cross just north of the chaotic terrain known as Conamara Chaos (splotch, a), are revealed to be sets of ridges bordered by dark smudges when viewed in favorable lighting at higher resolution (b). A very high resolution example (c), from the top of the center image, shows that ridges typically come in pairs, along either side of a crack. The largest ridge pair (c) is about 2 kilometers across and 100 meters high. Ridges may be produced by tidal processes (see Figure 4).

(Figure 3b). Now we see that the global-scale lines prove to be complexes of ridges, roughly parallel and somewhat intertwined. The dark lines are revealed to be simply a very diffuse darkening along the surface adjacent to the ridge complexes. Yet this diffuse darkening was the only indication of the lineaments when viewed at low resolution, where the ridges were too small to be resolved. The ridges, even the complexes of multiple ridges, lie entirely within the gap between the dark margins. In terms of morphology, the major crack systems are manifested by ridges.

The relation between cracks and ridges becomes clearer when we note that ridges come in pairs wherever we see them. A close-up of a densely ridged area just north of Conamara is a good example (Figure 3c). Because cracks are manifested as ridges, and ridges come in pairs, ridges probably are built along both sides of a crack.

Ridges may result from tides (Figure 4). Once a crack is created, it is worked on a daily basis as tides distort the icy crust. Suppose a crack reaches liquid water. Opening the crack slightly will allow liquid water to rush up to the float line, just as cracks on a frozen lake fill nearly to the surface with liquid. Where water is exposed to the surface, it must boil in the vacuum as it freezes in the cold. Within a few hours the top of the crack is filled with a few meters of new ice. But then the tides reverse, and the crack begins to close. The fresh ice is crushed. As the walls of the crack slam together, ice is squeezed to the surface. At the beginning of the next daily cycle, the crack opens again, leaving ridges along both sides. Quantitative estimates suggest that enough material to build ridges a kilometer wide and 100 meters high, typical of larger examples, could be extruded in this way in about 20,000 years.

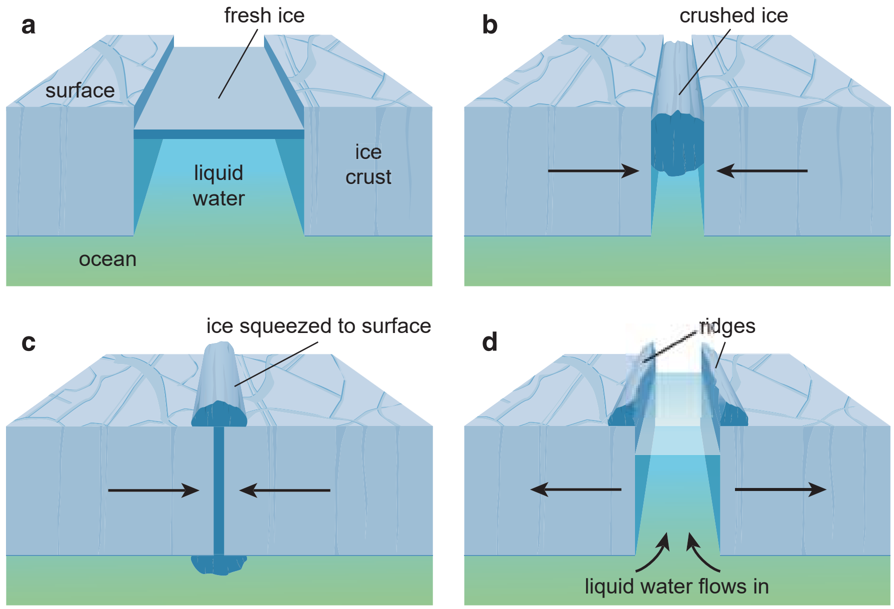


Figure 4. Ridges on Europa's surface may be formed over tens of thousands of years by the repeated opening and closing of cracks in the ice crust during the daily tidal cycle. At the beginning of the day (a), a crack opens, letting in liquid water, which freezes and boils at the surface, forming a thin layer of fresh ice. As the daily cycle progresses, the crack closes (b), crushing the ice and slush and squeezing some of it onto the surface of the crust (c). When the cycle begins anew (d), parallel ridges are formed on either side of the crack as it opens again the next day. In time, the steady accumulation of ice on the ridges may reach a height of 100 meters.

Cycloidal Crack Patterns

Global-scale lineaments are not the only indicators that tidal tension is the cause of cracks. Perhaps the best evidence comes from a distinctive and ubiquitous crack pattern in the shape of scalloped chains of arcs, in the geometric shape called cycloids. Figure 5 (left) shows an image of the southern hemisphere made by Voyager, in which these features were first seen, and some beautiful examples in the north are shown in Figure 5 (right). Numerous examples extend over 1,000 kilometers, including a dozen cycles or more, with each arc typically about 100 kilometers long.

These features were a mystery for nearly 20 years, until Randy Tufts and Greg Hoppa of the University of Arizona gave careful thought to the changes in tidal stress during the course of a Europan day. A crack begins when and where the tension exceeds the strength of the ice. Then, as the crack propagates across the surface, perhaps at a brisk walking speed of several kilometers per hour, the time of day advances and with it the strength and direction of the stress. The crack curves in response to the change in direction, and comes to a stop when the stress decreases below a critical threshold of strength. During the next few hours, while the stress is too weak to continue the cracking, its direction changes. Then, about a day after the cracking first began, the stress increases enough that the crack begins propagating again, in a new direction, leaving a cusp at the site where propagation was delayed. Thus each arc corresponds to one day's worth of crack propagation.

In quantitative terms, this model fits observed patterns well, but it requires a substantial tidal amplitude for the stress to overcome the strength of the ice. That amplitude can only be reached if there is a substantial liquid ocean. For that reason, the existence of the cycloids became the first convincing evidence that there is indeed a global ocean on Europa. Corroboration came from later Galileo fly-bys when the magnetometer instrument detected modulation of Jupiter's magnetic field, consistent with a conductive layer, such as a salty ocean, around Europa.

Strike-Slip Displacement

The surface of Europa is also modified by tectonic displacement of huge plates of surface ice. Consider the 1,000 kilometer-long dark band called Astypalaea Linea crossed by the more recent cycloids in the left-hand panel of Figure 5. A reprojected version, simulating a view from directly overhead, is also shown in Figure 6 (left). Geologist Tufts noticed two interesting characteristics of Astypalaea. Wispy white lines come up to the dark band from both sides and stop, and there are several parallelograms running along the length of the dark band. Both features are clues that Astypalaea is a strike-slip fault—there has been shear displacement of the opposite sides. When Tufts cut the image along the band and shifted the terrain on both sides, in effect running the shear backward in time, he found that the wispy lines reconnected and the parallelograms closed back up,

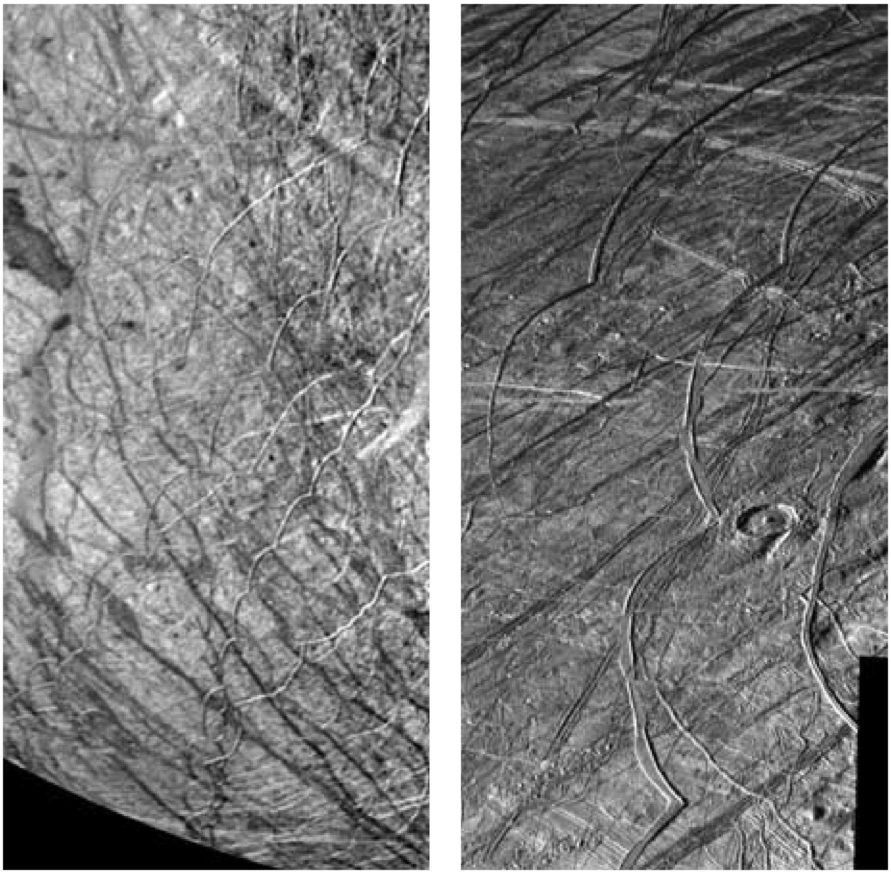


Figure 5. Cycloidal crack patterns are ubiquitous on Europa, including these in the southern (left) and northern (right) hemispheres. A typical chain of arcs may include more than a dozen cycles, with each arc about 100 kilometers long. Each arc corresponds to the propagation of a crack during a single day. (The southern hemisphere image was made by the Voyager 2 spacecraft, and is courtesy of the Jet Propulsion Laboratory and NASA.)

as shown in the right-hand panel of Figure 6.

Strike-slip displacement is common on the Earth, where moving plates of crust slide past one another. The San Andreas fault is a good example, and it is similar in size to Astypalaea. But what could drive such shearing movement on Europa?

Again the answer appears to be tides. When Tufts, Hoppa and I looked at the stress across Astypalaea during the course of a Europan day, we found that it cycled through a condition of tension that would gape the crack open, followed by shear, followed by compression that would slam the crack shut, followed by shear in the opposite direction. This cycle repeated itself daily. Although the shear stress reversed itself during the course of each day, the crack would have been open and easy to shear in one direction, and closed and hard to shear in the other direction.

This processes is analogous to walking, where we lift our foot from the floor, shear it forward, then press it against the ground and try to shear it back. Friction prevents the backsliding. In the same way, a fault on Europa can take daily steps, shearing the terrain on one side past the other side. This theory should be able to predict the direction of shear at any place on the satellite, and indeed there is a fairly good match between what is calculated and what is observed.

Strike-slip displacement by tidal walking requires that cracks penetrate all the way through the ice down to the ocean. Otherwise, the daily steps of the walking process could not occur. This result has profound implications because cracks cannot go very deep into Europa. More than a few hundred meters down, or maybe a few kilometers at most, the weight of the ice would squeeze so hard that tidal tension could not overcome it. If cracks cannot penetrate more than a few kilometers, but they go all the way down to the ocean, it means that the ice layer on Europa must be very thin.

Dilation

Large parts of Europa's crust have also shifted through the dilation of cracks. As with shear displacement, we can cut and paste the images, reclosing the cracks to restore the surface to an earlier configuration (Figure 7). The opened cracks are called bands, and they tend to display a characteristic morphology that must reflect the infilling of the opening by material from below. Most striking are the sets of fine-scale parallel ridges within most bands, often symmetrical about a central groove.

This structure, and its similarity to double ridges, suggests a way that dilation may be driven by tides. Suppose the process of ridge formation is not very efficient, in the sense that some of the newly frozen ice within a crack does not get completely extruded during the daily squeezing phase. The crack may not be able to close completely. One day at a time, the crack may be forced open by increasing amounts of jammed material. Even if some other forces, such as currents in the underlying ocean, act to pull plates apart, the daily tidal working would have prevented a smooth, uniform rate of opening and may have been responsible for the fine-scale parallel lines that formed during dilation.

One other way that tides have driven dilation is by strike-slip displacements. If one large sheet of crust is moving past another, there may be a place at the end of the shear zone where the crust is pulled apart. Similarly, if strike-slip displacement occurs along a crack that is not straight, pull-apart zones may be created as part of the necessary geometry for accommodation. The large parallelogram that opened in the strike-slip fault Astypalaea appears to have formed in this way. What's more, high-resolution images of Astypalaea show that it probably started as a cycloidal crack, and as it sheared a chain of pull-apart zones were created, each with the multiple striations around a central groove that are characteristic of typical dilational bands.

Dilation has been considerable on Europa, and as one looks back through the time sequence represented by cross-cutting and dissected features, it becomes evident that such extension of the surface has been going on for a long time. Given this source of new land, where is the sink? One possible explanation may come from the ubiquitous chaotic terrain.

Chaos!

Nearly half of the surface of Europa is made of chaotic terrain, visible as dark splotches in full-disk views of the moon. Its detailed character becomes evident in a sequence of images (Figures 3a, 3b and 8), which progressively zoom in on the archetypal example, Conamara Chaos. Here the surface has been disrupted, with rafts of older terrain displaced within a lumpy matrix.

Given the evidence from tectonics that the ice crust is thin, one plausible explanation for the chaotic terrain is melt-through, perhaps all the way from the ocean below. After the larger rafts of surviving crust move about, the water between them refreezes (filled with smaller lumps), leaving the characteristic appearance of chaotic terrain. Tidal heat, generated by internal friction, could be enough to keep the ice thin, and only modest concentrations are needed to melt patches of ice every now and then. Undersea volcanoes could provide such sources of concentrated heat.

Early descriptions of Galileo images reported that patches of chaos-type terrain had a characteristic size of about 10 kilometers across. This observation was interpreted to mean that these features were manifestations of the tops of convection cells in viscous solid ice at least 20 kilometers thick. That model of chaos formation was very different from the melt-through model, and completely inconsistent with the thin ice inferred from tidal-tectonic theory.

However, study of more recent images shows that the 10-kilometer size is an artifact of observational bias introduced by the limited imaging data available. In fact, patches of chaotic terrain can be more than a thousand kilometers across, and recent statistics show that smaller patches occur in increasing numbers as their size decreases. For many of the Galileo images the limits of resolution prevent us from detecting patches of chaos smaller than 10 kilometers across. So the lower limit of the observed size depends directly on the resolution of the images. This observational bias created the false impression that most patches were 10 kilometers across.

While the creation of chaotic terrain destroys older surfaces, chaotic terrain is itself destroyed by tectonic processes. In Figure 8, for example, we see a couple of cracks and incipient ridges wending through the rafts across Conamara Chaos. The history of Europa has been an ongoing interplay of resurfacing, by tectonics and by chaos formation, with each destroying what was there before and with each evidently involving a breakthrough of the ocean to the surface.

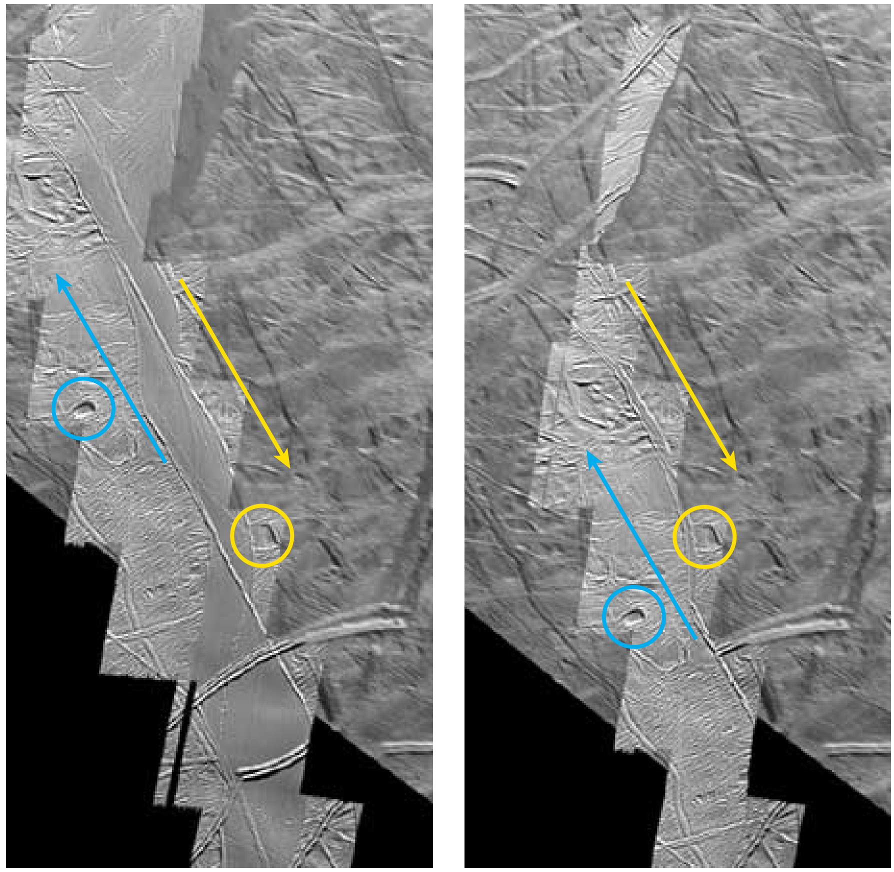


Figure 6. Strike-slip displacement (arrows) of Europa's crust is evident in the Astypalaea region by comparing its current appearance (left) with a reconstructed image of its past appearance (right). The relative positions of landmarks (such as the circled features) on either side of the fault line reveal the extent of the motion. The tidal mechanism that created the fault suggests that Europa's crust must be relatively thin, perhaps only a few kilometers deep. The inserts (lighter shade) display higher-resolution images of the region, which spans a distance of about 200 kilometers from top to bottom. (The mosaic at left was created by the University of Arizona's Planetary Image Research Lab. The reconstruction at right is based on the work of Randy Tufts.)

The continuous creation of so many openings may hold the key to the puzzle of the disappearing surface. Large sheets of surface can be readily compressed without leaving any signs of compressional stress if they are full of holes. In this way chaos may provide the sink that accommodates tectonic dilation.

Tides Drive Rotation

Tidal theory suggests that a satellite on a circular orbit will quickly come to rotate synchronously, with the same period as its orbit, much as Earth's moon does (so that one hemisphere always faces us). But, because of Europa's eccentric orbit, the Jovian-induced tides will maintain a spin rate slightly faster than synchronous so that Europa's face toward Jupiter gradually changes.

This means that the tidal stress experienced by any given piece of real estate undergoes gradual changes as a consequence of the non-synchronous rotation. A crack that is actively worked, building ridges, may later freeze shut. A walking strike-slip fault may stop advancing and freeze in place. A cycloidal crack pattern may retain a record of the daily stress where it was formed, but now be much further east than where that happened.

In fact, cycloids can be used to determine the rotation rate. From tidal theory we can determine the longitude of a cycloid's formation and the order in which the cycloids were formed over the geological history of the surface. We can place a limit on the duration of that history from the paucity of impact craters. By combining such information, Greg Hoppa has inferred a rotation rate for Europa that suggests that the hemisphere currently facing Jupiter was last in this position about 50,000

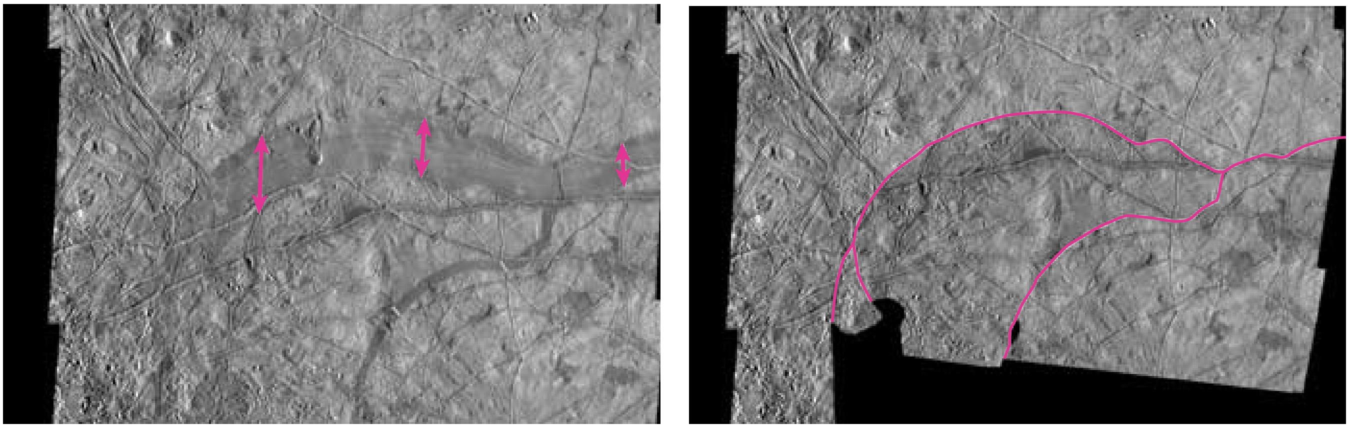


Figure 7. Dilation of cracks in Europa's crust forms bands that appear to be filled with material from below (left). The appearance of a crack before the dilation (right) can be reconstructed by cutting and pasting parts of the original image. Tidal processes can explain the existence of such dilation features. These images cover a region about 350 kilometers across. (Reconstruction by Randy Tufts, University of Arizona.)

years ago.

At any given location on Europa, tidal tectonic processes are regular over thousands of years, but over tens of thousands of years, local conditions change in important ways.

Life?

The emerging picture of Europa as a world with an ocean that is intimately linked to its surface describes a physical setting that may provide everything needed for life. In contrast, the notion of a very thick ice layer that isolates the ocean from the surface provides a less hospitable setting. In that picture, the ecosystem is isolated from both oxygen and from sunlight. Scientists attracted to the possibility of life on Europa have been forced to imagine alternative biochemistries, assuming volcanism and hypothetical metabolisms. Even with the freedom to model deep-sea conditions unconstrained by any observations, there has been concern that life would be very limited, should it exist at all.

The thin-ice picture that follows from the ideas of tidal tectonics overcomes these problems. Consider a crack in the ice that is actively worked, opening and closing on a daily basis (Figure 9). At the base is liquid water, just above the freezing point, containing a mixture of substances from the moon's interior and from external sources, such as comets. These substances leave the orange-brown traces visible wherever the ocean reaches the surface, whether through linear cracks or chaotic melt-through.

The surface of the ice is bombarded by energetic, charged particles from Jupiter's magnetosphere, creating oxidants (such as oxygen and hydrogen peroxide), which get mixed back into the ice. Cometary material lands on the surface, depositing its suite of organic and other substances. Organisms within a few centimeters of the surface would be killed by the radiation, but enough sunlight could penetrate a few meters below to drive photosynthesis.

As a crack opens and closes, relatively warm seawater flows up and down each day. Much like tidal zones on Earth, this niche could conceivably support a rich ecology. Plants might anchor in the sunlight near the surface. Other organisms might grab, tick-like, onto the walls of the cracks and tap the passing daily flow as it mixes the disequilibrium chemistry. Some of them might break loose as ice melts beneath their feet, or be covered occasionally by new ice on the walls. Still others, floating like jellyfish, might simply go with the daily flow from ocean to surface.

Such a niche might be stable for thousands of years. But as Europa rotates relative to Jupiter, the crack moves

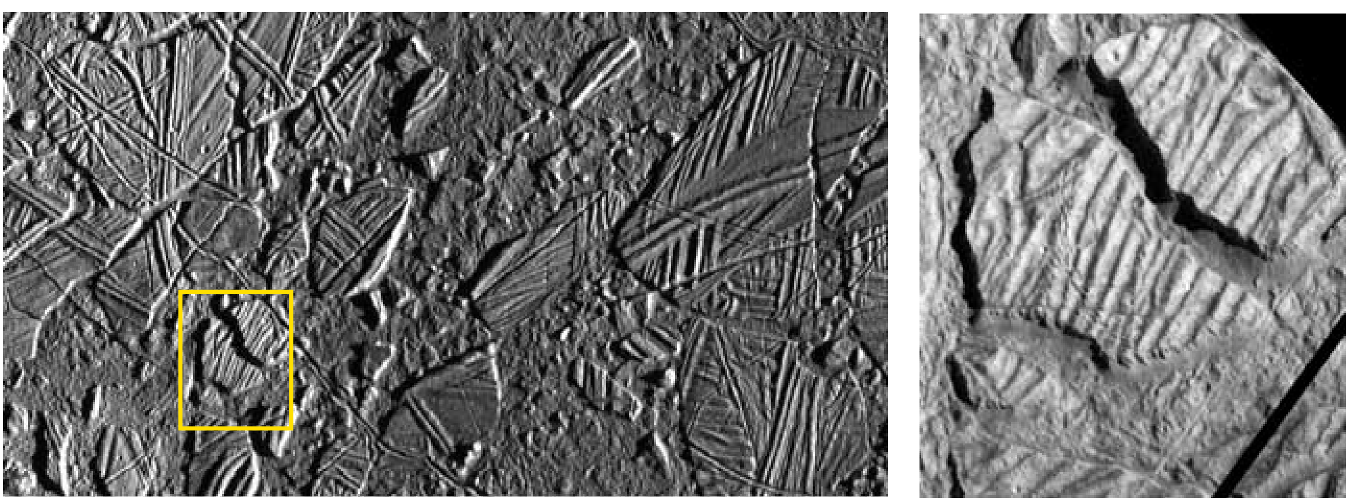


Figure 8. Chaotic terrain, here in Conamara Chaos, appears as regions of disrupted crust, perhaps where the surface ice has melted. Rafts of surviving crust appear to have moved through the melt and are redistributed within a lumpy matrix of newly frozen ice. Internal friction may keep Europa's crust thin, and undersea volcanoes may provide concentrated sources of heat where the crust melts through to the ocean.

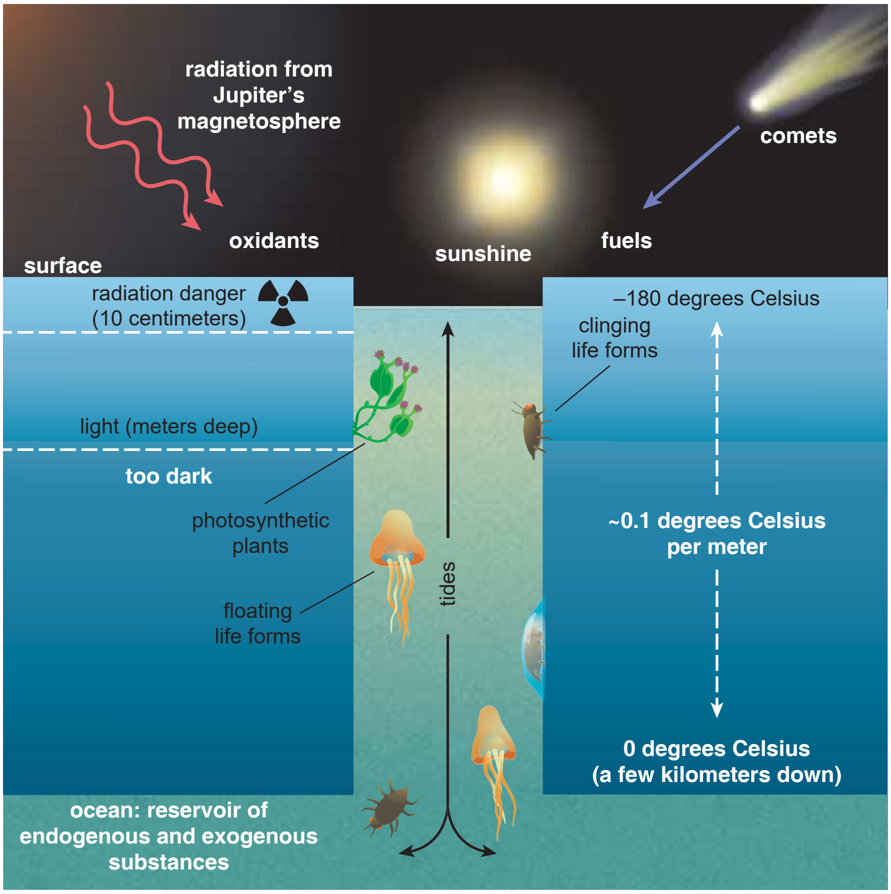


Figure 9. Europa's biosphere may include several forms of life that have adapted to niches provided by the cracks in its thin icy crust. Although radiation from Jupiter's magnetosphere would pose a danger to life on the surface of Europa (to a depth of a few centimeters), sunlight could sustain photosynthetic organisms (flower icon) beneath this layer to a depth of several meters. Clinging life forms (bug icon) might scale the walls of the cracks (and perhaps hibernate within the walls), whereas floating life forms (jellyfish icon) may be able to float with the tides as the cracks open and close on their daily cycle. The life-form icons used here do not imply a particular shape or appearance of the life that might have evolved on Europa. Earth-style flowers, bugs and jellyfish are unlikely on Europa, but their analogues may exist in such a setting.

to a different stress regime. The daily working might cease, sealing the crack closed and freezing organisms within it. For life to go on, some organisms must escape and travel through the sea to an active crack. Or the creatures frozen in the crack might be able to hibernate until a later thaw. They would only have to wait a million years or so, a feat demonstrated by Antarctic bacteria on Earth. By that time, a chaos-forming crustal melt event is likely. Even sooner, new cracks might cross the area, releasing organisms into a new home.

As long as individual niches remain stable, they would allow organisms to be comfortable, secure and prosperous, but the longer-term change due to rotation would drive adaptation and mobility. These challenges are an important requirement for driving evolution to more complex and diverse forms of life.

Not only would the tidal tectonic processes provide habitable settings in the crust, they may also allow life to exist and prosper in the ocean by providing access to oxidants. Oceanic life would likely be part of the same ecosystem as organisms in the crust.

If there has been life on Europa, it is likely to be there now and to be readily accessible. The youth of Europa's surface tells us that the physical processes and conditions that potentially allow for life on Europa have been in effect during the last one percent of the age of the solar system. Because they were so recent, they likely continue today. Moreover, should there be a biosphere on Europa, it may extend from deep in the ocean up to within a few centimeters of the surface.

This possibility makes Europa an exciting target for future exploration. Extraterrestrial life may be more accessible than previously thought. Rather than needing to drill down through many kilometers of ice, we may be able to scoop up organisms at or near the surface. What would make exploration easier for us is not necessarily good for the Europans, however. If the near-surface is as fecund as now seems plausible, Jupiter's moon may be vulnerable to contamination by terrestrial hitchhikers on our spacecraft. Explorations need to be planned with care.

Even if Europa proves to be sterile, the complex suite of geophysical processes and their unique relationships with geological and dynamical phenomena make Europa one of the most active and exciting bodies in the solar system.

Bibliography

Chyba, C. F., and C. B. Phillips. 2001. Possible ecosystems and the search for life on Europa. Science 98:801–804.

Greenberg, R., P. Geissler, B. R. Tufts and G. V. Hoppa. 2000. Habitability of Europa's crust: The role of tidal-tectonic processes. Journal of Geophysical Research 105(E7):17,551–17,561.

Greenberg, R., G. V. Hoppa, B. R. Tufts, P. Geissler, J. Riley and S. Kadel. 1999. Chaos on Europa. Icarus 141:263–286.

Greenberg, R., and B. R. Tufts. 2001. Infecting other worlds. American Scientist 89:296–299.

Hoppa, G. V., B. R. Tufts, R. Greenberg and P. E. Geissler. 1999. Formation of cycloidal features on Europa. Science 285:1899–1902.

Peale, S. J., P. Cassen and R. T. Reynolds. 1979. Melting of Io by tidal dissipation. Science 203:892–894.

Reynolds, R. T., S. W. Squyres, D. S. Coburn and C. P. McKay. 1983. On the habitability of Europa. Icarus 56:246–254.

Schenk, P., and W. B. McKinnon. 1989. Fault offsets and lateral crustal movement on Europa: Evidence for a mobile ice shell. Icarus 79:75–100.

Tufts, B. R., R. Greenberg, G. Hoppa and P. Geissler. 1999. Astypalaea Linea: A San Andreas-sized strike-slip fault on Europa. Icarus 141:53–64.

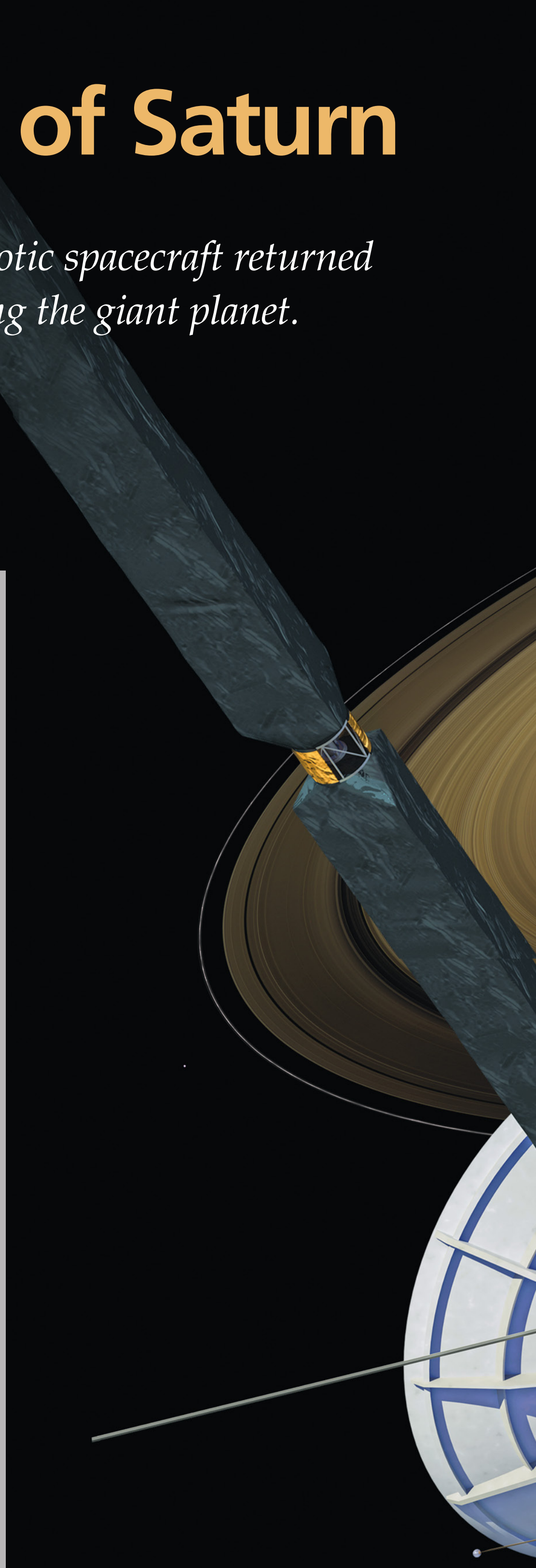
Links to Internet resources for "Tides and the Biosphere of Europa" are available on the American Scientist Web site:

http://www.americanscientist.org/articles/02articles/greenberg.html

Cassini and the Rings of Saturn

As one of its final scientific feats, the accomplished robotic spacecraft returned unprecedentedly detailed data on the material encircling the giant planet.

Matthew S. Tiscareno



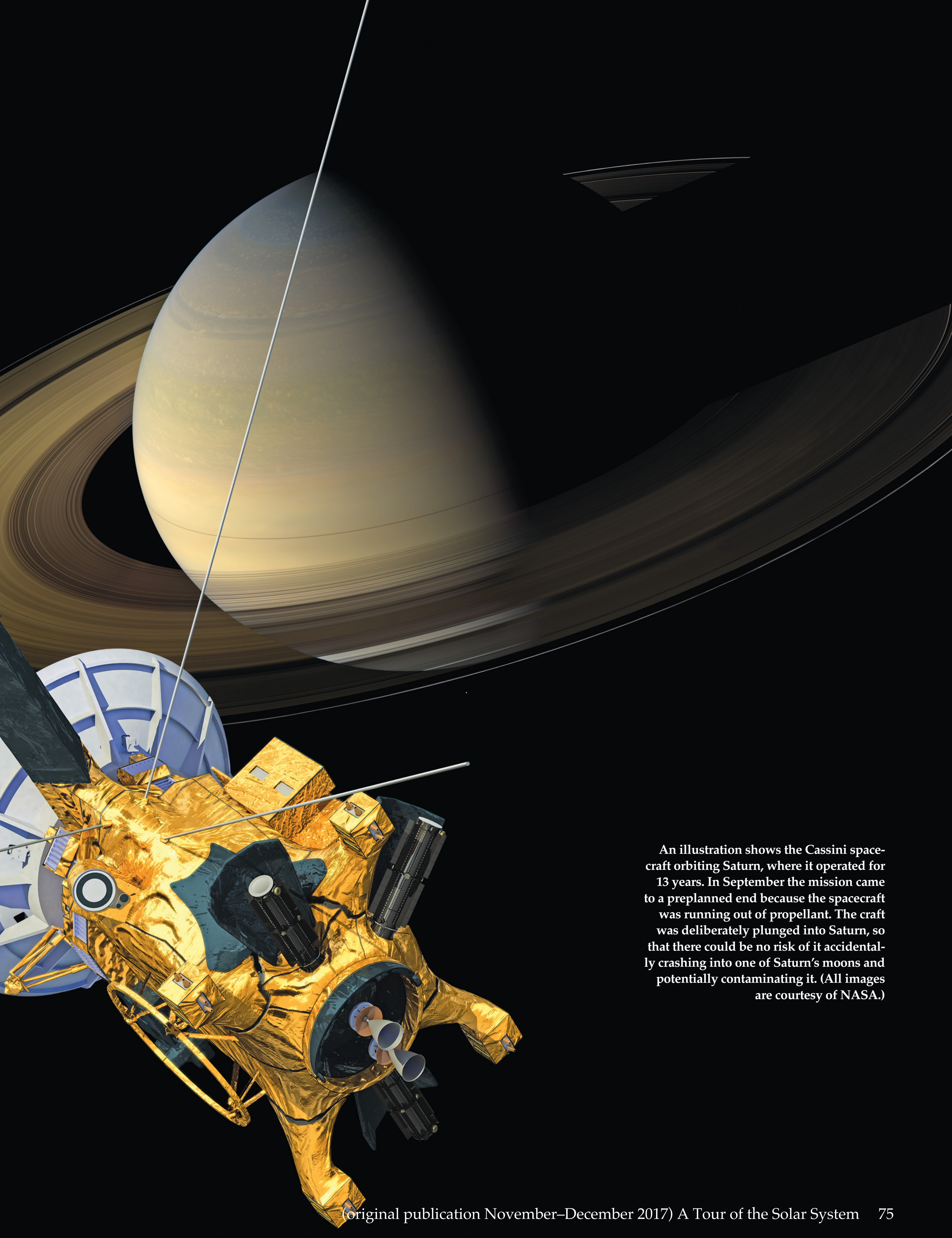
The rings of Saturn are delicate, enigmatic, beautiful, and useful. NASA's Cassini spacecraft has been studying them (and the rest of the Saturn system) over the course of a spectacularly successful 13-year mission, but its concluding year has been the most spectacular of all—practically a whole new mission. In December 2016, Cassini commenced weekly plunges through the ring plane just off the outer edge of the main rings (activity termed the "Ring-Grazing Orbits"), and transitioned in April 2017 to weekly plunges between the rings and the planet's cloud-tops (the "Grand Finale" orbits). The mission concluded in September 2017 with a final descent into the planet's atmosphere, to preclude possible terrestrial contamination of Saturn's moons.

The science goals of the Ring-Grazing Orbits and the Grand Finale orbits included direct sampling of particles from Saturn's rings and atmosphere, detailed measurements of Saturn's gravity and magnetic field to probe the planet's interior, and unprecedentedly close-range imaging of Saturn and its rings with the spacecraft's main camera and similar instruments. These observations followed up on discoveries made through the prime and extended Cassini missions, continued monitoring seasonal or evolving phenomena, and took advantage of Cassini's unique proximity to Saturn in these closing stages of the mission.

Saturn's planetary ring system is made up of countless chunks of ice, each in its own orbit around the planet Saturn. The chunks range from marble-size to house-size. The ring system is confined to the plane of Saturn's equator and is arguably the flattest structure known to humanity, with an end-to-end dimension equivalent to circling the Earth seven times, but a vertical height about that of a house.

Saturn's rings are useful to scientists because they provide physical and chemical clues regarding the formation and history of the entire Saturn system, they serve as detectors and amplifiers for planetary phenomena around them, and they help us understand more generally how disk systems operate, providing clues about other kinds of disks, such as baby solar systems. A few of the most compelling science questions regarding Saturn's rings are: How do ring particles interact with one another, with moons embedded within them, and with moons farther away? Are ring particles made of ice that is fluffy or dense, pristine or sooty? What are their shapes and sizes? What structures do we see in the dusty parts of the rings, and what can we continue to learn from them?

Matthew S. Tiscareno is a senior research scientist at the SETI Institute in Mountain View, California, and a participating scientist and imaging team associate for the Cassini spacecraft. He received his PhD in planetary science from the University of Arizona in 2004. His research focuses on the orbital and rotational motions of rings, satellites, and planets. He is the coeditor of Planetary Ring Systems *(Cambridge University Press, 2018). Email: matt@seti.org*



An illustration shows the Cassini spacecraft orbiting Saturn, where it operated for 13 years. In September the mission came to a preplanned end because the spacecraft was running out of propellant. The craft was deliberately plunged into Saturn, so that there could be no risk of it accidentally crashing into one of Saturn's moons and potentially contaminating it. (All images are courtesy of NASA.)

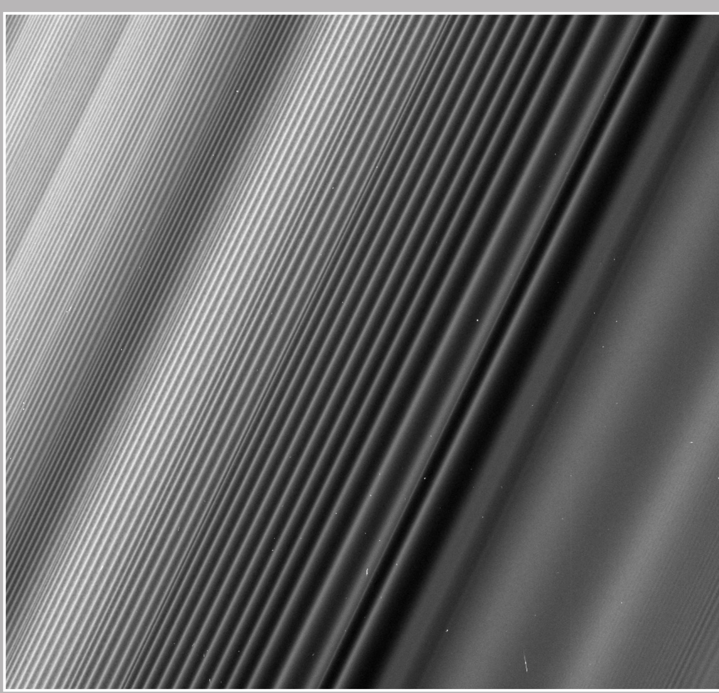
The structure of Saturn's ring system is dominated by concentric bands and tightly wound spirals. Very few of these bands are empty gaps, so it is better to think of the entire system as a disk, rather than as "countless rings." Much of the structure is not well understood by scientists, but the most common type of understood structure are waves that propagate through the rings at locations where ring particle orbits *resonate* (or essentially "hum in tune") with the orbit of a moon orbiting beyond the rings. Cassini scientists discovered in recent years that a few of these waves are excited by structures within the planet Saturn.

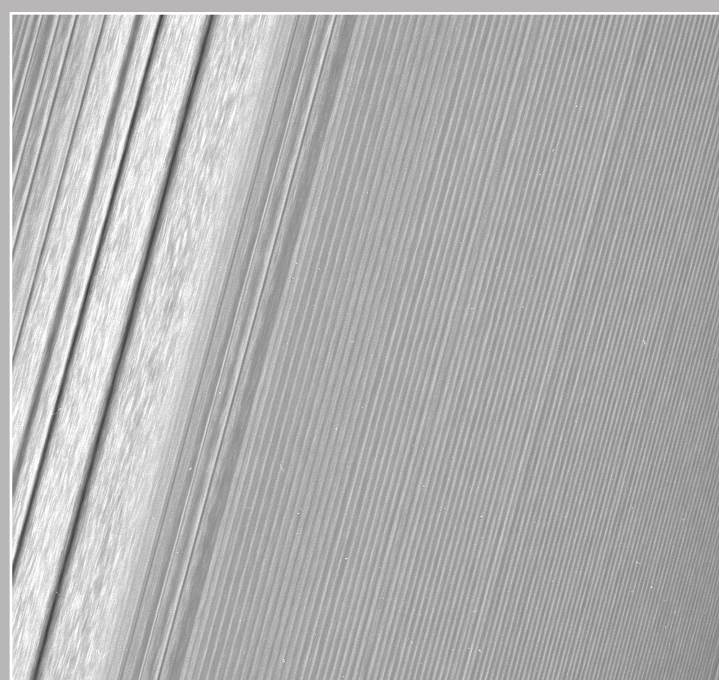
Moons that orbit within the rings deflect nearby ring particles with their gravity and thus create a disturbance around themselves. The moons Pan and Daphnis are large enough that this disturbance becomes a sharp-edged empty gap extending around the entire circumference of the rings. Extrapolating from this archetype, scientists expected each of the dozen other sharp-edged gaps in the rings to also host a moon apiece; however, careful searches by Cassini have shown that this is not the case, so these gaps must have a more complex origin. Moons in a smaller size class than Pan and Daphnis create a propeller-shaped disturbance that is filled back in by the rings before it can extend circumferentially.

The interior and surface properties of ring particles are poorly known. Are they more like ice cubes or snowballs? Are their surfaces fluffy or slushy or frosty? The answer may be all of the above. Images from the Ring-Grazing Orbits and Grand Finale have shed new light on the collective interactions of ring particles. A speckled strawlike texture occurs within waves at locations where the particles had been recently compressed but that pressure has now been released, perhaps as a result of temporary clumps falling apart. More surprisingly, sharp-edged bands of similarly speckled texture occur at many locations within the rings, as well as streaky texture and other types of texture. Each of these textures likely indicates a different mode of mutual interaction among the ring particles, and/or different ring particle properties.

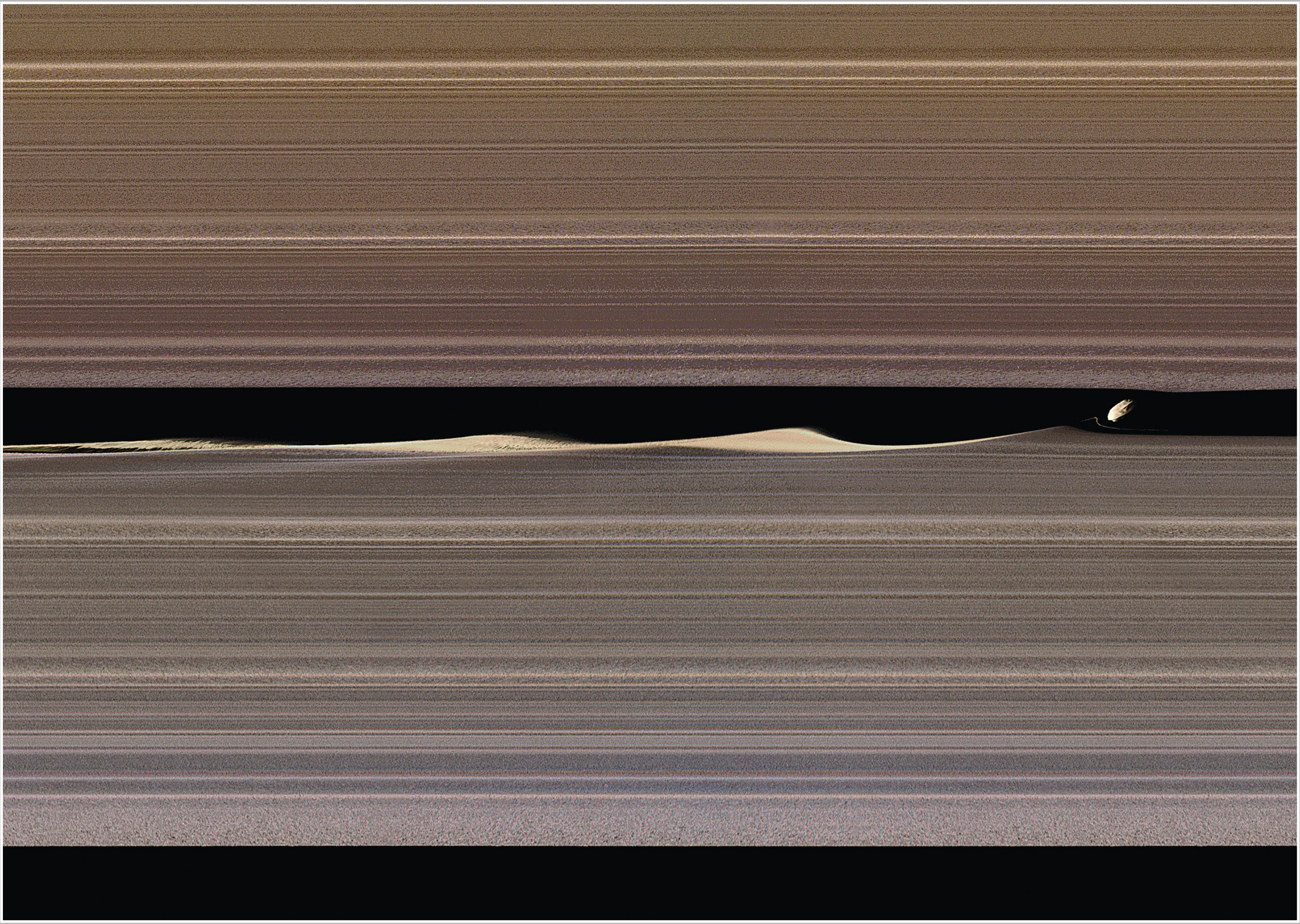
The main rings are largely free of dust (that is, of icy particles that are smaller than the period at the end of this sentence), because dust tends to collect as a layer on the surfaces of larger ring particles. In the few places where dust does coexist with larger particles, it indicates the presence of vigorous ongoing activity that is keeping the dust in motion. In locations that are far from the main rings, only dust is present, and the dust then can be sculpted into shapes that reflect the forces surrounding it.

Perhaps the most fundamental mystery surrounding Saturn's ring system is its age. Certain aspects of the dynamical interactions between the rings and moons can only be rewound for about 100 million years, indicating either that a significant reworking of the ring system's structure occurred within that time, or that the entire ring system is only that old. An important clue in this mystery is the rate at which the rings are being polluted by "soot" falling in from the Solar System. If that infalling rate is low, or if the rings have a relatively large mass of water ice with which to mix the pollution, then it would be reasonable to conclude that the rings are as old as Saturn and still sport the relatively pure water ice we see them to have. However, if the infalling rate of pollution is high and the mass of the rings is low, then the rings are likely no older than 100 million years. Cassini's dust detectors were working on a definitive measurement of the pollution rate, and the mass of the rings will be measured directly from the gravitational pull of the rings on Cassini during the Grand Finale orbits. After analysis of the final Cassini data has been completed, these questions should find more concrete answers.

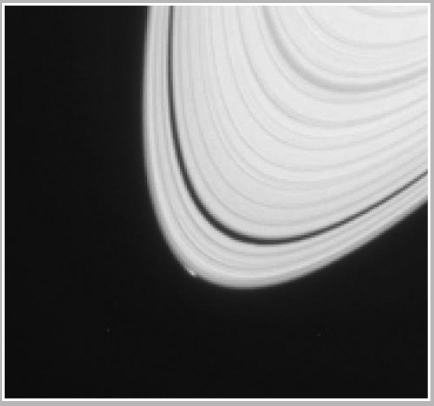




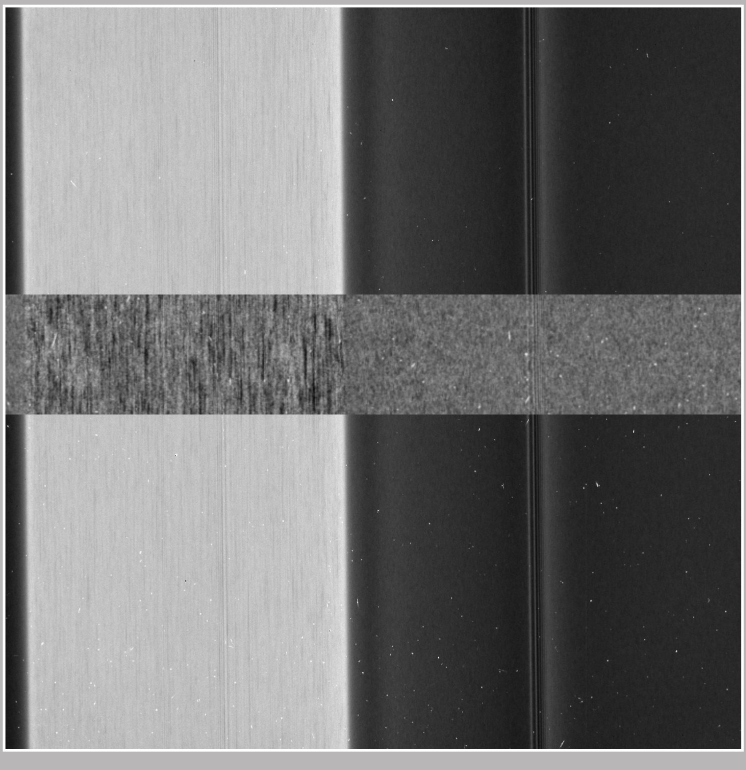
The *spiral density wave* at left occurs at the location where ring particles orbit Saturn twice for every time that the moon Janus orbits once (called a *2-to-1 resonance*). Because Janus trades orbits with its partner moon Epimetheus every four years, this wave sports several "glitches" that trace the wave's history going back several decades. The spiral density wave at right is also caused by the gravitational pull of the moon Janus, but at a location where ring particles orbit Saturn four times for every three Janus orbits (a *4-to-3 resonance*). The peaks of the wave are dark, and the broad, bright troughs contain a speckly, strawlike structure, likely due to transient clumps of ring particles that are now falling apart. The corduroy-like structure at center and right is a result of a recent pass of the moon Pan.



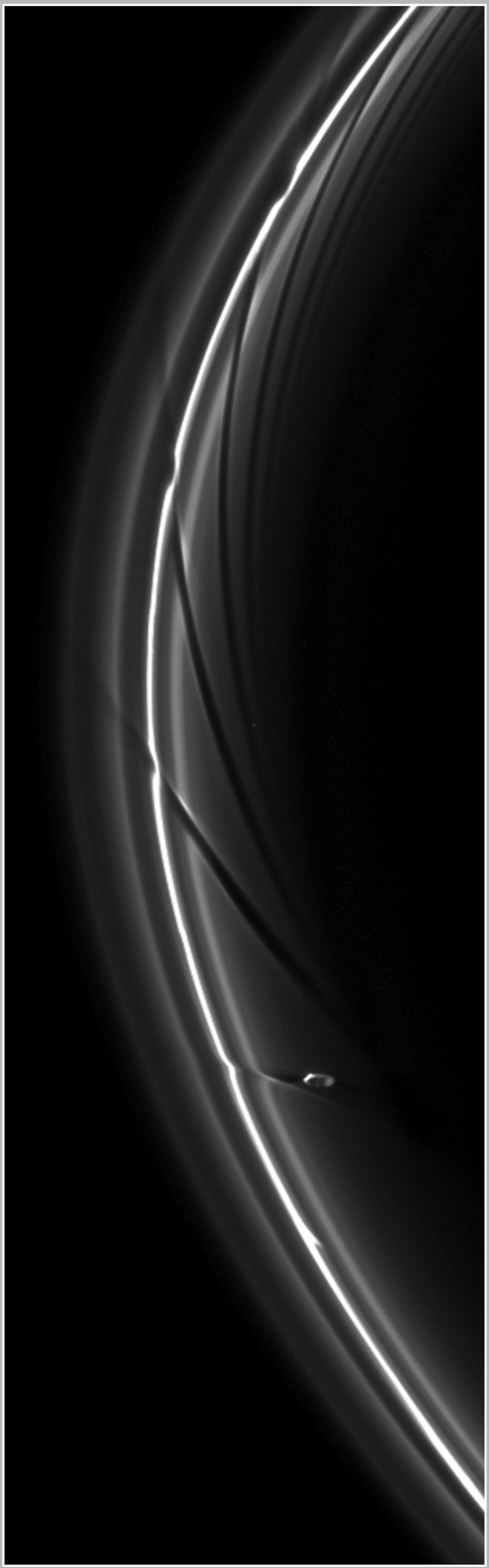
The moon Daphnis is seen orbiting within its sharp-edged gap. Ring particles that have recently passed the moon are perturbed into wavelike structures in the gap edge, but Daphnis's influence is waning by the third wavecrest back, which is disintegrating into a collection of small clumps.



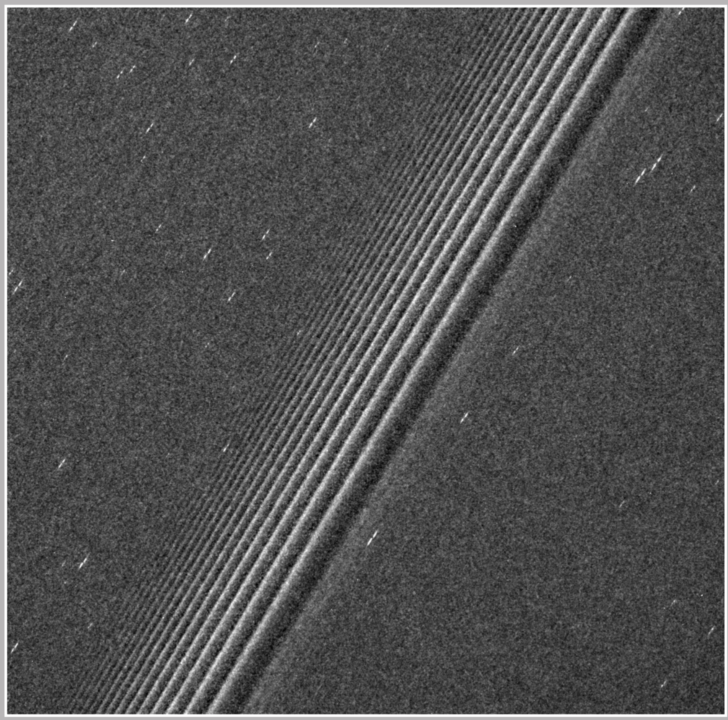
The quasi-moon "Peggy" appears at the outer edge of the main rings, looking as though it might spiral away from the rings and become a freely orbiting moon. Subsequent Cassini images have shown Peggy's story to be more complicated, because it has broken into several pieces and has remained within the rings.



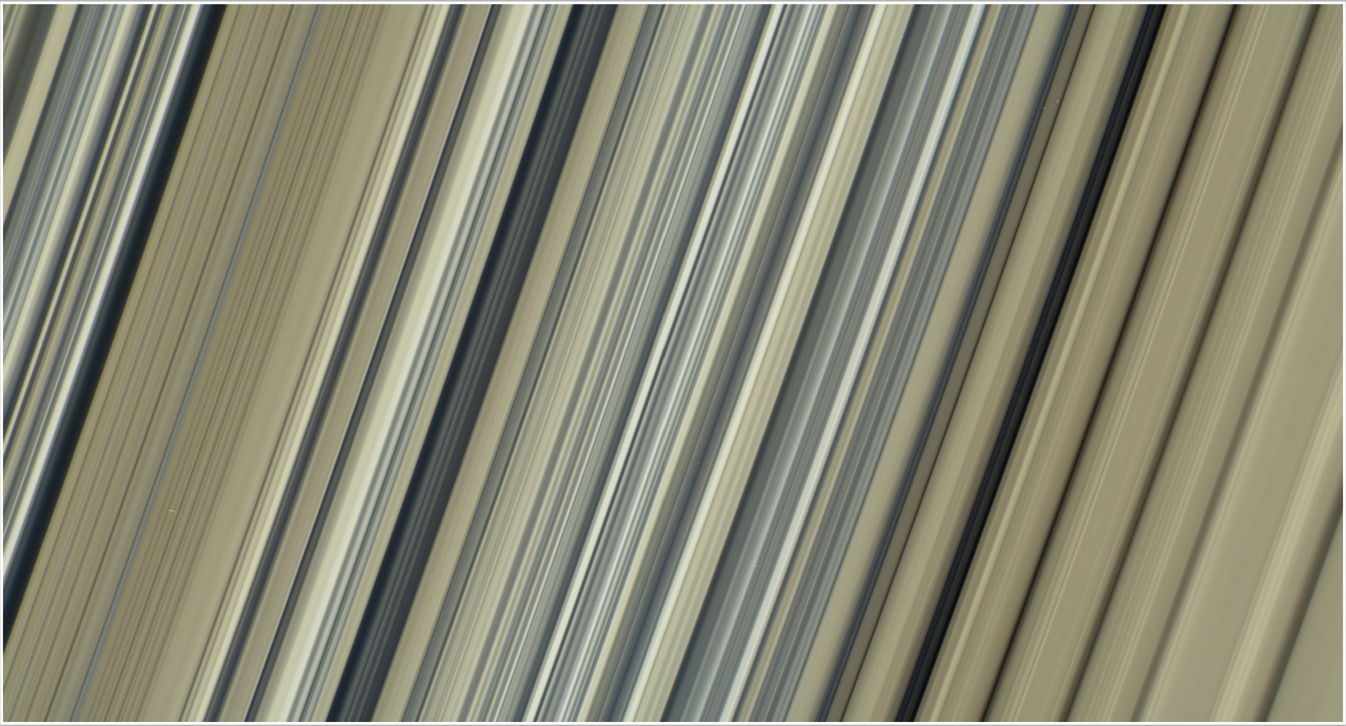
At left, a sharp-edged bright band (called a plateau) has an unknown cause. The central strip, from which the first-order brightness has been removed so that local structure becomes more visible, shows that the plateau has a streaky texture whereas the surrounding region has a speckly texture. The different textures must indicate different modes of particle properties and/or interactions. Also visible to the right in this image is a spiral density wave driven by structure inside the planet Saturn.



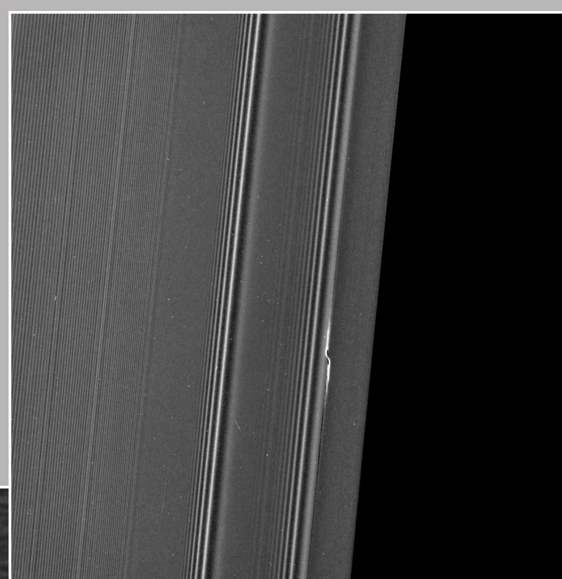
The moon Prometheus dips into the dusty F ring and sculpts its material. The resulting channels flatten out as they move downstream from Prometheus. Also visible in this image are several strands of dust comprising the F ring, with some bright knots that indicate embedded moons.



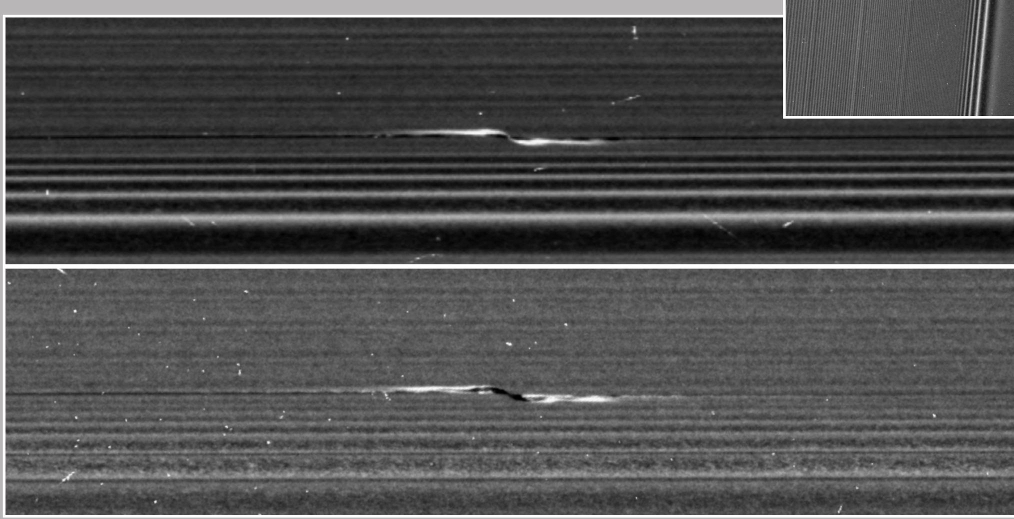
Swarms of small *propellers* of debris (*above*) surround a density wave (with a 9-to-8 resonance with the moon Prometheus). Each propeller contains a 100-meter moonlet at its core, which creates the disturbance around itself.

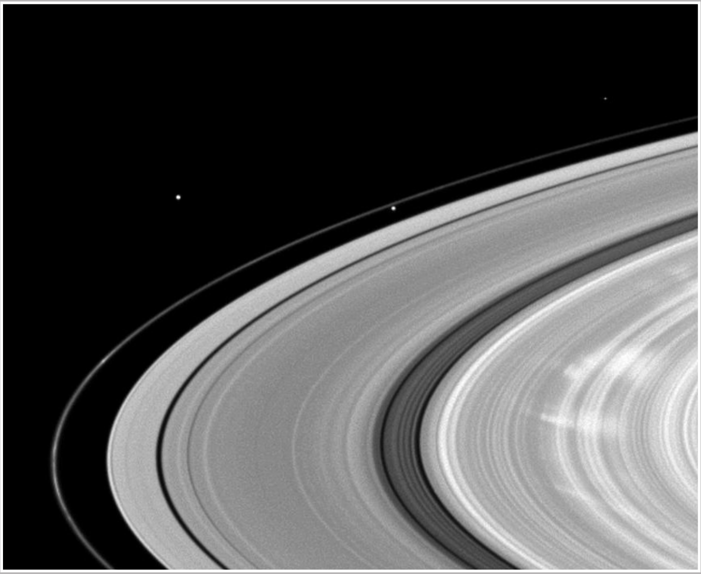


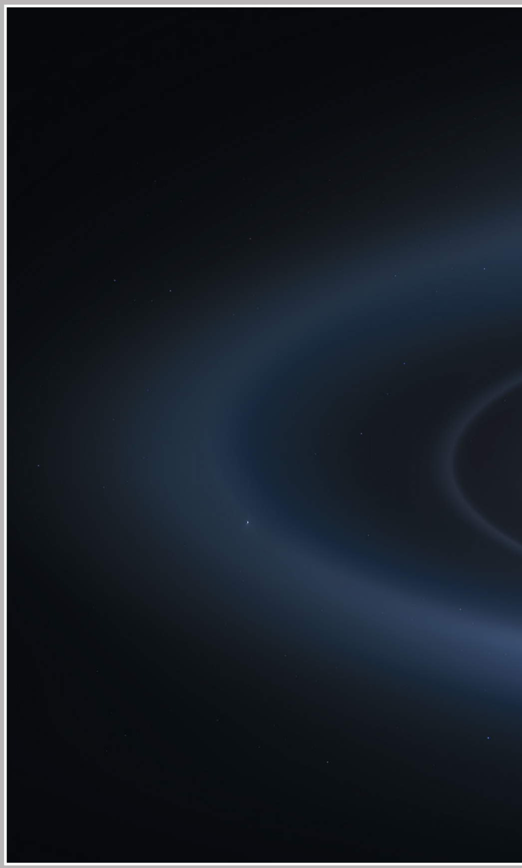
An inner-central part of the planet's B ring is shown in what is the highest-resolution true-color image of any part of Saturn's rings produced to date. The reason for the tan color is poorly understood, as is the origin of the sharply defined banded structure.

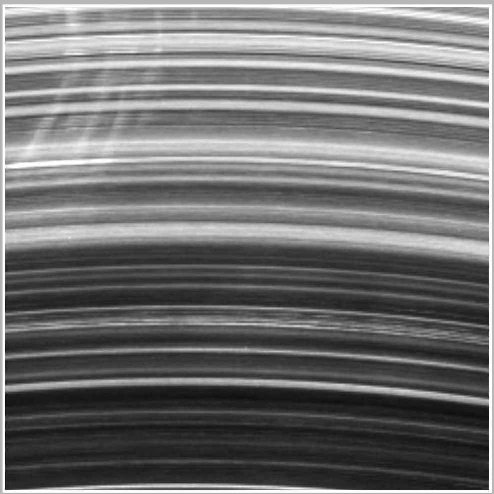


Close-up views show giant *propellers* caused by ring-embedded moons. The figures immediately below and to the right show the lit side of the rings, where brighter tones indicate more material reflecting light. The figures at far bottom and on the facing page bottom show the translucent, unlit side of the rings, where most dark tones indicate more material, such that the ring becomes opaque. Cassini has tracked the orbits of these individual propellers for more than a decade and has found subtle changes in the orbits that are likely due to moon-disk interactions. These individual propellers have been nicknamed "Santos-Dumont" (*below*), "Earhart" (*right*), and "Blériot" (*facing*).

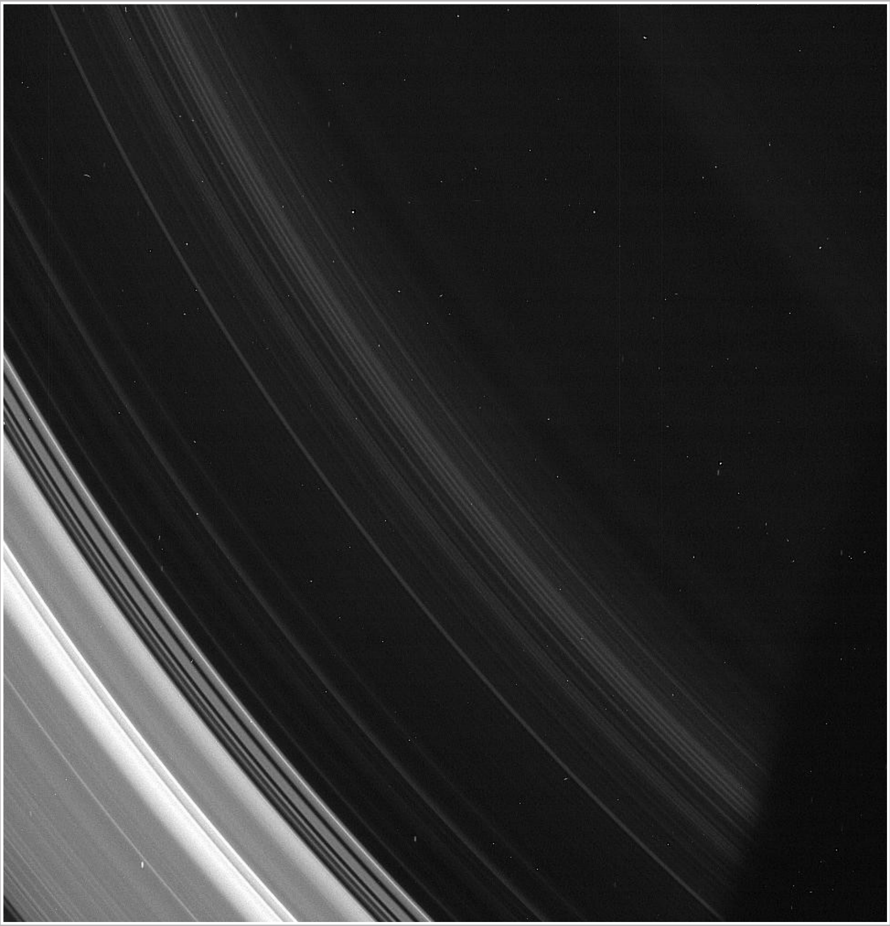




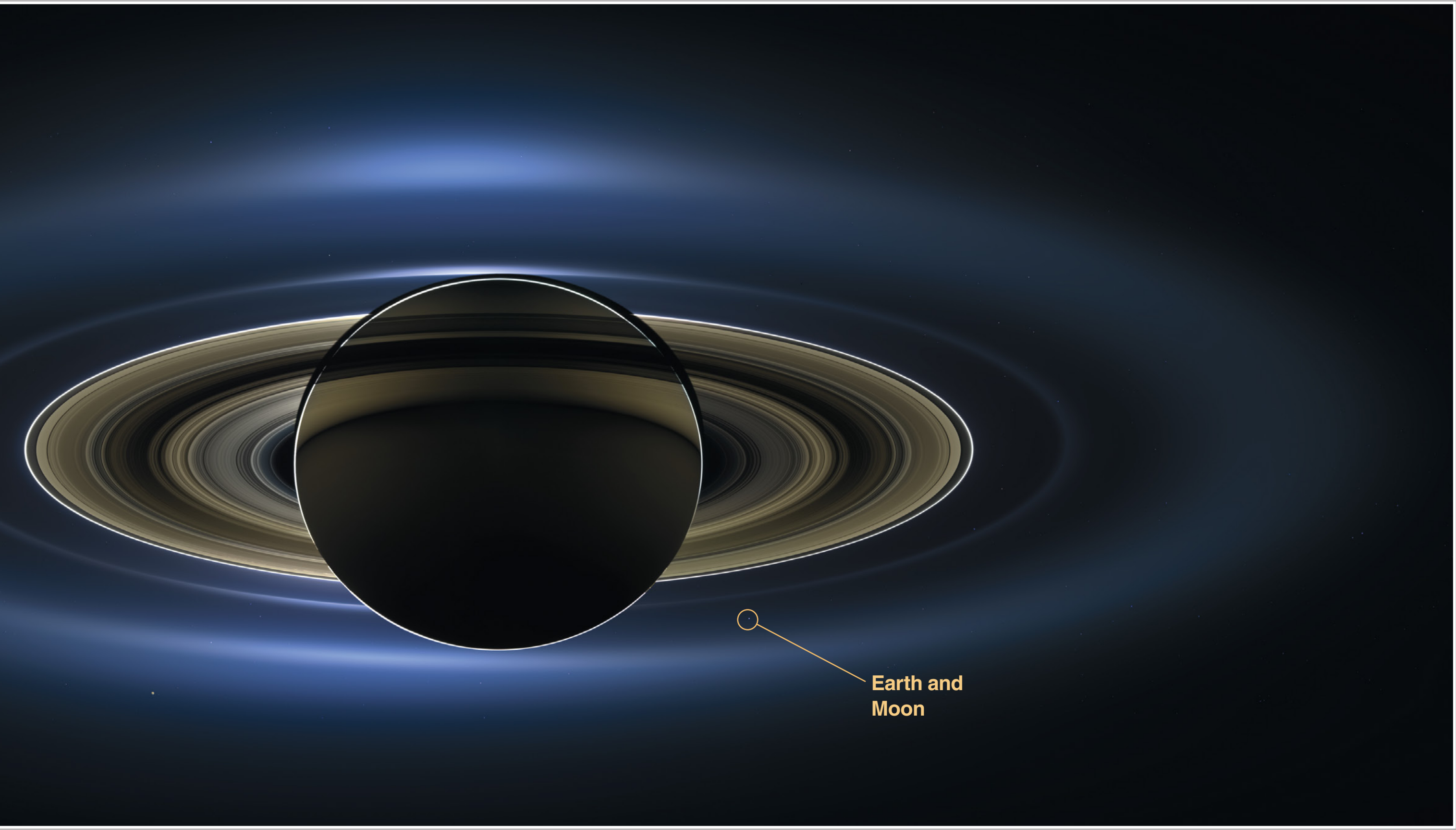




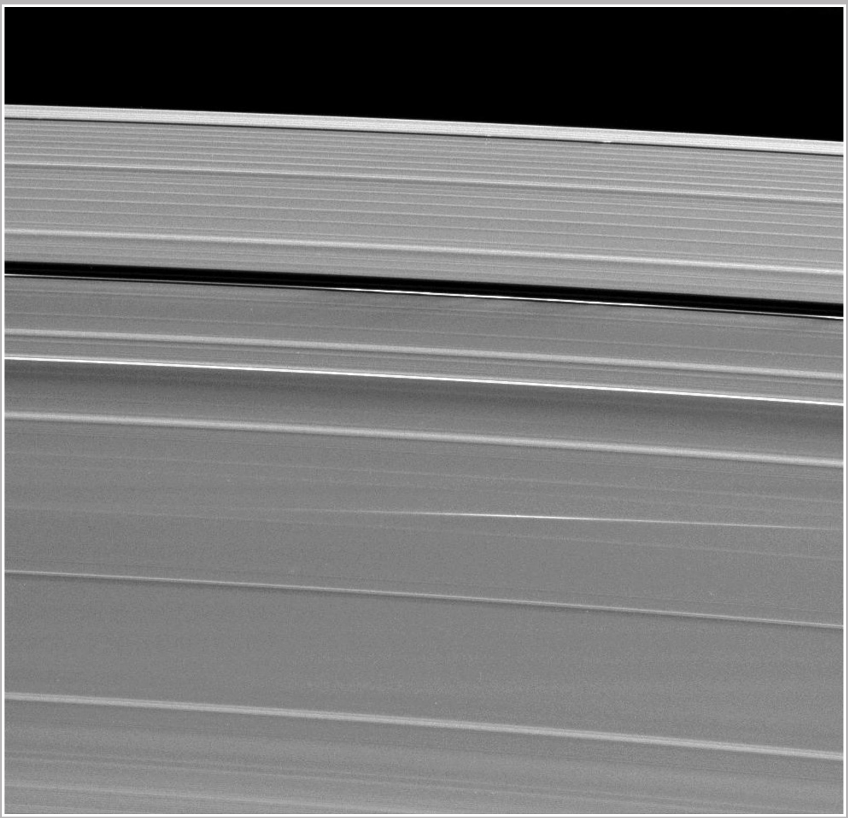
Ghostly white markings called *spokes* appear in the central region of the main rings, in a wide-angle image (*above*) and in a more detailed image (*left*). These spokes are composed of dust levitating above the ring plane. The dust was likely ejected by meteoroid impacts and then caught up into Saturn's magnetic field, although the details are not well understood.



Dusty lanes occur within the D ring, the innermost ring of Saturn. Some of the bands comprise a spiral that preserves traces of an inferred meteoroid impact that took place in the 1980s. The brighter bands at bottom-left are the innermost portion of the main rings.



A panoramic view of Saturn shows both its main rings and its outer, dusty rings. Because the Sun was behind Saturn when this image was taken (that is, Cassini was inside Saturn's shadow), the dusty rings appear nearly as bright as the main rings, and dusty regions that are more closely aligned with the Sun appear even brighter. The bluish E ring is composed of material ejected from the geysers on the south pole of the moon Enceladus. The Earth and its Moon are visible to the lower-right of Saturn.



An impact ejecta cloud tells the tale of a meter-sized object that plowed through the rings some hours before this image was taken. The ejecta cloud is near the center of the image and is nearly horizontal, canted at an angle from the concentric bands of the ring structure.

For relevant Web links, consult this issue of *American Scientist Online:*

www.amsci.org/magazine/issues/2017/november-december

Improve Your Image

Were planetary scientists scooped by a chat group of amateur enthusiasts?

"Image is everything," goes the advertising adage. A curious twist of that notion emerged following the landing in January of the European Space Agency's (ESA) probe Huygens on Titan, Saturn's largest moon. The first images beamed from the probe via the Cassini spacecraft thrilled scientists but barely inspired most of the public. Peter Hartlaub, writing in the *San Francisco Chronicle*, lamented:

> While children once huddled in front of their radios and television sets, waiting for the latest updates on the fates of heroes such as John Glenn and Neil Armstrong, modern space missions all seem to end the same way: with indistinct pictures of orange rocks, followed by impassioned hyperbole from scientist types attempting to convince us how totally awesome the images are.

Yet, thanks to a few amateur image analysts, "awesome" images of Titan were available on the Internet within hours of the release of the raw data.

How did this happen? Were the scientists "scooped?" Well, yes and no. It turns out that scientists had meant to release the raw data but, according to ESA Huygens project manager, Jean-Pierre Lebreton, not quite so quickly. Apparently the University of Arizona server onto which the information from Huygens's cameras was uploaded was "made accessible [to the public] by mistake." The Huygens data were therefore available before the scientists had a chance even to look them over. And the public was waiting. Not the public whom, Hartlaub writes, consider "space exploration… really boring," but a loose-knit cadre of space-imaging enthusiasts who convene via Internet chat rooms and who showcase pictures on personal Web pages. One of these Web sites is run by Anthony Liekens, a doctoral student in biomedical imaging at the University of Eindhoven in the Netherlands.

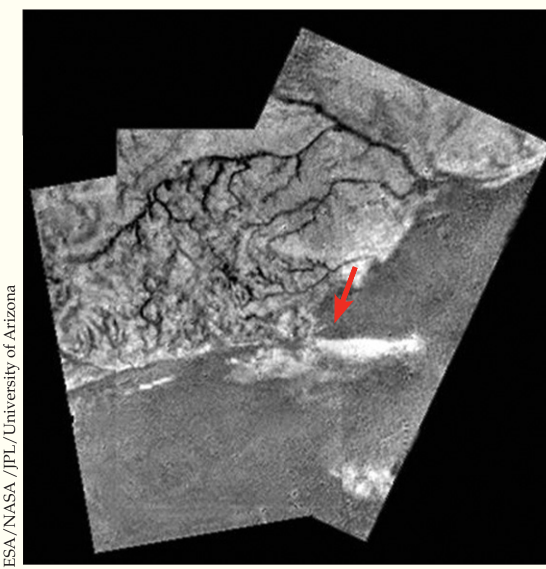
Liekens's chat group was anticipating the Huygens data after they'd read a University of Arizona announcement that raw pictures would be available soon after the landing. But events unfolded much faster than Liekens expected. Within hours, his amateur group had used standard image-manipulation software, such as Photoshop and Terragen, to render ESA's low-resolution grayscale composites into serene landscapes complete with coastline, clouds and islands—scenery recognizable to earthlings.

Were the amateurs doing science, or just prettying up the pictures? Perhaps a bit of both. While the rendered photos *were* attractive, interpreting the raw images and using software to create realistic views of Titan required a sophisticated understanding of image analysis and some knowledge of planetary science. As the Huygens probe parachuted down to the surface, the Descent Imager Spectral Radiometer (DISR) designed by scientists at the University of Arizona captured about 350 triplet images, using three cameras at different angles and magnifications. These low-resolution images overlap to create larger mosaics that look rather like aerial photographs. After compiling mosaics of Titan's surface from the triplets, the amateurs converted these from two-dimensional monochrome to three-dimensional color. However, as Liekens himself pointed out, the professionals are best equipped to render the most realistic views, as they have the expertise to interpret nonvisual data that may provide clues to features not evident from the DISR pictures.

In any case, the Titan landscapes ended up widely distributed across the Internet, although, Liekens noted, "The big media outlets like CNN and BBC didn't pick it up right away." The pictures first made their way to "nerd Web sites and blogs" and then filtered across cyberspace to the media giants.

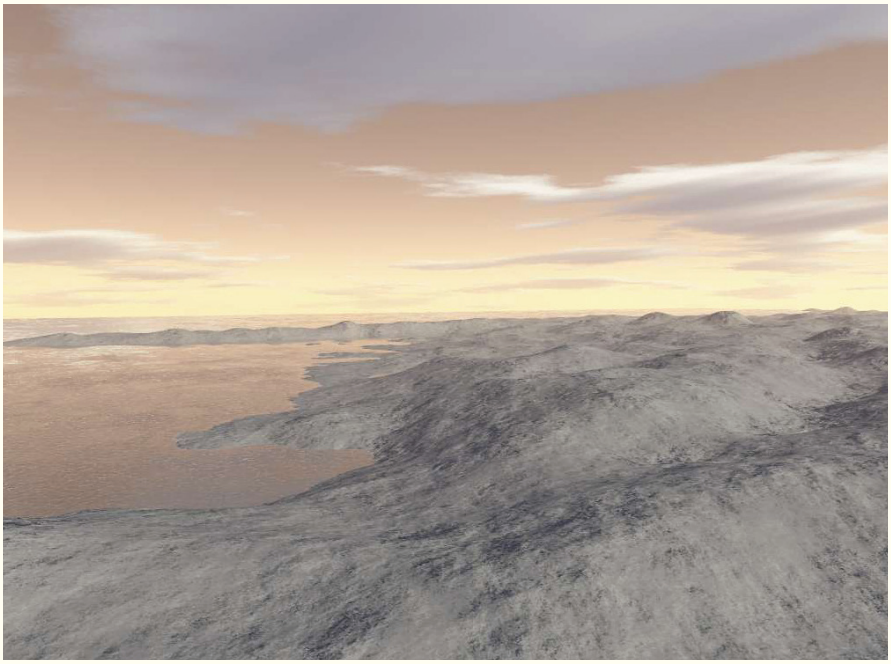
The rapidity and scope of the images' distribution gave the strong impression that the amateurs had beaten scientists to the punch. Lebreton says his team was impressed: "Our scientists here [at ESA] looked at the images and said, 'Wow, they're beautiful.' Their beauty was not matched by the images we released."

There seem to be no hard feelings. Lebreton says the amateurs should be given credit for clearly stating up front that their embellishments were not necessarily accurate but meant to be enjoyed for what they represented. In fact, he says ESA is looking to hire some of the amateurs who worked on the images. Lebreton thinks that the release of the raw data has been an unexpected public-relations success. ESA officials will meet shortly to discuss the implications of the unintended experiment for future public relations.



ESA/NASA/JPL/University of Arizona

About 10 kilometers above Titan's surface, Huygens's imaging instrument revealed startlingly Earth-like topographic features, including rivers and a coastline. The arrow indicates the approximate viewpoint amateurs used to construct Titan scenery.



ESA/NASA/JPL/University of Arizona/Mike Zawistowski

Amateur image analysts rendered raw data from the Cassini-Huygens probe into three dimensions to provide an impression of Titan's surface. Created with the scenery-generating software Terragen, the topography is based on actual data, but the coloring is a guess.

The implications go far beyond the ESA offices. These events suggest amateurs are poised to contribute in significant and unexpected ways. New technology and access to cheap computing promise to dramatically change the amateur-professional interaction.

Although technology has moved astronomy into the category of "big science," amateurs still participate extensively. Astronomy has long been a favorite discipline of amateur scientists (among whom birding is the most popular), and amateur astronomers are often the first to detect comets or supernovae. For this reason, professional astronomers tend to value the contribution of amateurs. "Generally, the attitude [toward amateurs] is positive," says Robert Milkey, executive officer of the American Astronomical Society, "Many professionals are eager to collaborate and want to write amateur contributors into research projects and proposals.… The distinction between professionals and the best-qualified amateurs is that the professional is paid."

Ed Flaspoehler, president of the American Amateur Astronomical Association, who has worked as a data-processing consultant for major corporations, perceives the professional community differently, as "generally diffident toward amateurs."

"Some professionals view amateurs as a pool of graduate students who do the work while [the professionals] get the credit," he says. "Professionals have not yet figured how to put this [resource] to good use."

Even if only a small proportion of the estimated 250,000 amateurs in the U.S. want involvement at a higher level, Flaspoehler noted, that's still a large number compared to the six or seven thousand professionals. New approaches, such as those demonstrated by Liekens's chat group, make it possible for greater numbers of amateurs to be involved in more sophisticated ways.

Milkey thinks that "the technology [now] coming into the hands of amateurs is capable of doing serious science." Flaspoehler sees the Huygens imaging chat group as the latest trend in an evolutionary process. Early amateurs made visual observations, then graduated to telescopes. In the 1980s, amateurs widely adopted film as a recording medium. During the 1990s CCD (charge-coupled device) technology became widespread among observatories and then became affordable to amateurs in the form of digital cameras. Over the past few years amateurs have adapted digital cameras to taking astronomical photographs and therefore became involved in image processing.

Some scientists see that the spread of new, cheap technology, particularly in information and data management, offers new ways to interact with the public. One example is so-called distributed computing. Internet-connected personal computers, while otherwise idle, perform subsets of calculations that require massive processing power. Starting with SETI@home, which analyzes radio telescope data to detect extraterrestrial signals, the use of distributed computing has extended to processing data from laser interferometers to detect gravitational waves (Einstein@home) and running models of global climate-change prediction (climate*prediction*.net) and protein folding (Folding@home).

"There's a real need for organization to bring amateurs and professionals together," Robert Milkey remarked. He hopes that members of the public will see more of what keeps scientists coming to the observatory or laboratory every day—and have the chance to participate in scientific research and experience that excitement for themselves.—*Roger Harris*

The Centers of Planets

In laboratories and computers, shocked and squeezed matter turns metallic, coughs up diamonds and reveals Earth's white-hot center

Sandro Scandolo and Raymond Jeanloz

Those who enjoy following planets in the night sky got a special treat this summer as Mars passed nearer to Earth than it had come in 60,000 years. Even at close range, though, the Red Planet retained an air of quiet mystery. All the planets have it: In contrast to the twinkling stars around them or the lights of the noisy city, the planets appear peaceful and immobile. And in fact planets *are* mysterious; in some ways we know little more about them than did the ancients who worshiped them. In particular, we have few clues as to what any of the planets (even our own) are like on the inside.

What we do know is that the interior of a planet is not a peaceful place. From the evidence that exists, we can infer that the interiors of planets typically are subjected to pressures more than a million times that of Earth's atmosphere at the surface and that temperatures in their centers reach several thousand degrees Celsius. You can think of each planet's interior as a giant foundry specialized for processing a particular chemical composite under extreme conditions. These composites range from the simple hydrogen-helium mixture of Jupiter and Saturn to the more complex mixture of "ices" (water, ammonia and methane) that compose Neptune and Uranus, and finally to the mostly "solid" internal structures (silicates plus iron in solid and sometimes liquid form) of terrestrial planets such as Mars, Venus and of course Earth. "Solid" here is a bit of a stretch; over geological time planetary-scale objects made of rock, metal or ice deform and exhibit convection just as fluids do. Likewise the substances we call ices are not strictly solid; they exist as gases in the outer atmospheres of giant planets and as fluids in the interior.

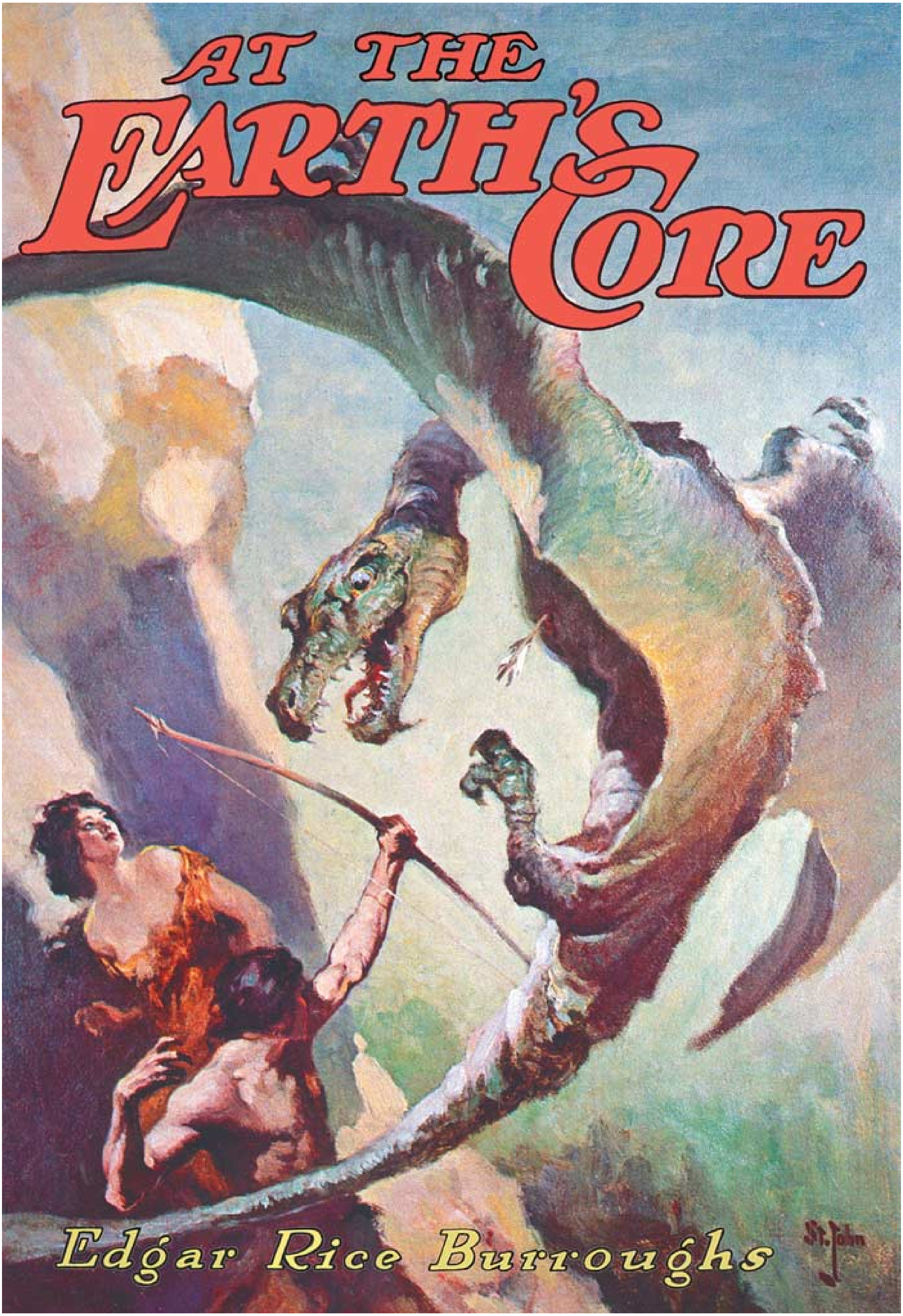
The interiors of planets are totally inaccessible, so what we know comes from indirect measurements and analysis. For example, seismic waves detected at the Earth's surface tell us a great deal about the internal structure of our planet. Similarly, measurements of mass, gravitational moments (variations in the strength of gravity at different positions above and around a planet), magnetic fields and a few other quantities, taken by space probes or remote observation, allow us to infer the density profiles and internal dynamics of all the planets of the solar system. Estimating pressure is a fairly straightforward matter because we have reliable equations to calculate pressure from mass and depth—the same equations that tell a deep-sea diver how fast pressure will increase during a descent. Surface observations—for example, the chemical makeup and thickness of the atmosphere—can shed further light on the composition of a planet.

Unfortunately the information one gets is only enough to make crude estimates. And it is hard to imagine a probe capable of penetrating the skin of a planet to a depth of more than a few miles and bringing back a sample of material from the interior. In its 1996 encounter with Jupiter, the Galileo probe made a successful 600-kilometer-deep dive into the giant planet, revealing unexpected features of the outer layers. But 600 kilometers is a scratch on the surface of Jupiter, whose radius is 70,000 kilometers. The deepest anyone has ever drilled into the Earth is 12 kilometers, just 0.2 percent of the distance to the center. And there is every reason to expect that the samples from such limited probes may not be representative of the planetary interior.

Frustrated by the lack of concerted effort to send probes into the deeper regions of the Earth, David Stevenson at the California Institute of Technology recently made a "modest proposal." He

Sandro Scandolo was recently appointed senior staff member at the Abdus Salam International Centre for Theoretical Physics (ICTP) in Trieste, Italy, where his research encompasses simulations of high-pressure phase transitions in covalent, molecular and metallic systems in addition to surface science, polymers and nonlinear optics. He received his Ph.D. in physics from the Scuola Normale Superiore in Pisa in 1993. He then moved to the International School for Advanced Studies (SISSA) in Trieste, where he became assistant professor in 1998 and associate professor in 2002. He spent a two-year sabbatical leave (2000–02) at the Princeton Materials Institute, Princeton University. Raymond Jeanloz is professor of Earth and planetary science and of astronomy at the University of California, Berkeley, where his group studies the nature and evolution of planetary interiors, as well as the properties of materials at high pressures. Address for Scandolo: Abdus Salam ICTP, Strada Costiera 11, I-34014 Trieste, Italy. Internet: scandolo@ictp.trieste.it

Figure 1. What lies at the center of Earth? The public imagination has freely explored the places that scientists cannot. Edgar Rice Burroughs, best known as the father of Tarzan, was one of many authors who have created imaginary worlds beneath our feet. This book jacket imagines life in Pellucidar, a "world at the Earth's core" that figured in seven novels written between 1913 and 1944. In Burroughs's conception, the Earth's crust is only 500 miles thick, leaving a vast hollow interior accessible via an opening near the North Pole. This "savagery of unspoiled Nature" is inhabited by dinosaurs, huge mammals and a variety of intelligent native races. A rather different picture of Earth's core emerges from experiment and external measurements, but much remains unknown about the centers of our planet and others.



Cover restoration by Phil Normand / Normand Design & Restoration

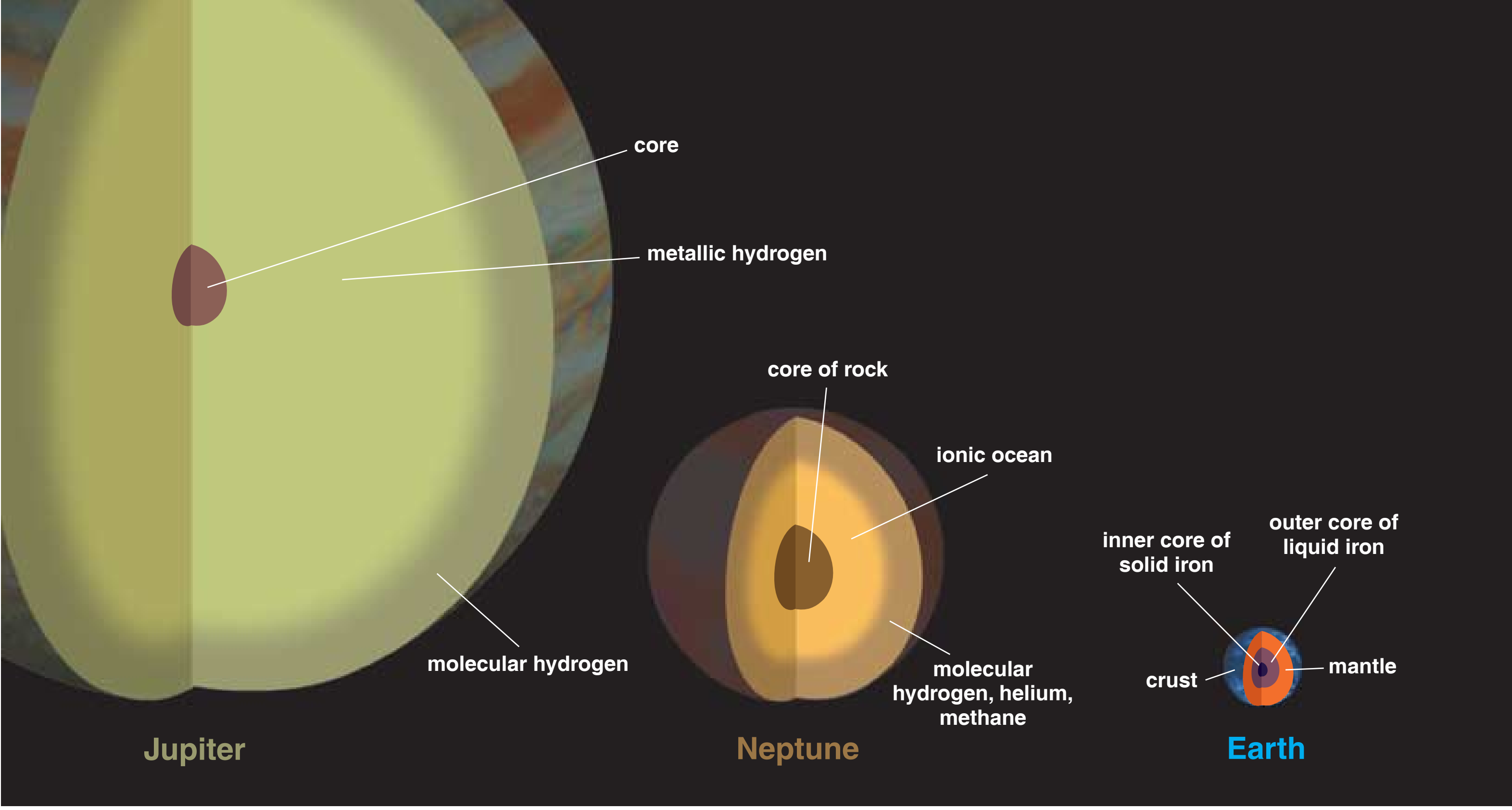


Figure 2. Jupiter and Saturn, the solar system's gas giants, are known to be composed of the simplest element, hydrogen, mixed with some helium, but recent experiments have confirmed that under high-pressure conditions hydrogen becomes a metallic fluid. Jupiter's center is thought to contain a core of rock at extreme pressures. The composition of Uranus and Neptune is richer, including water, ammonia and methane. Experiments and simulations in this case suggest that these molecules dissociate, creating an ionic ocean between the gaseous outer layer and solid core. Finally, Earth holds membership in the terrestrial planets, where a mantle of silicate and oxide rock gives way to a mostly iron core that, in the case of Earth, has a solid inner core surrounded by a liquid outer core.

claims that with a fraction of the financial outlay required to launch a space mission, a million tons of liquid iron could be poured into an artificial fracture at the Earth's surface. The iron would slowly but inexorably dive toward the Earth's center, and it could carry along insoluble probes that would send first-hand information from the bowels of the planet. With his tongue-in-cheek suggestion, Stevenson has captured the frustrations of geophysicists eager for ways to plumb the deep mysteries of planets.

Stevenson is also not the first scientist to reach for creative solutions. In fact, a large international community is exploring a completely different approach to the study of planetary interiors. Instead of trying to gain direct access to Earth's inner workings, some scientists have been striving since the early 20th century to simulate the conditions of pressure and temperature that shape planetary interiors. Microworlds created in the laboratory open a spectacular window into the composition, dynamics and evolution of planets, and may even offer a glance into the history of the solar system and how it evolved to its present form.

It is not easy to produce pressures of a million atmospheres and temperatures of a few thousand degrees inside the walls of a laboratory, let alone sustain them in a controlled way to allow sufficient time for measurement. Fortunately experiments can be complemented by theoretical calculations grounded in quantum and statistical mechanics, which can simulate from first principles the conditions existing deep inside planets.

Diamonds Are Not Forever

In the dark rooms of the Geophysical Laboratory at the Carnegie Institution of Washington, Dave Mao and Russell Hemley are getting closer each day to being able to reproduce in a controlled way the extreme conditions found in planetary interiors. Mao and Hemley are leaders in the use of cells that use diamond anvils to create extreme pressures (see "The Diamond-Anvil Cell," May–June 1992). As the hardest known material, diamond is well suited to the task of squeezing substances to a few million times atmospheric pressure.

To carry out the job, a pair of brilliant-cut gems, each usually weighing about one-quarter of a carat, is embedded in a powerful press. Unfortunately, the higher the pressure exerted by the pistons and screws, the greater the chance that one of the two diamond anvils that compress the sample—each typically a few millimeters in diameter—will fail, causing the experiment to implode suddenly with a single loud blast.

Mao reckons that he has broken hundreds of diamonds—small ones, fortunately. But the dismay caused by the occasional failure of a diamond is more than matched by the thrill of the amazing discoveries that have been made possible by this tiny device. When diamonds do withstand the load to which they are subjected, the pressure that can be reached at the center of the anvil's tip, a spot a few tens of micrometers large, is enough to reproduce the conditions found along a considerable fraction of a planetary radius.

Squeezing matter to planetary pressures dramatically alters its macroscopic properties, including some that are essential in planetary modeling: density, mechanical strength, viscosity and electrical conductivity. Substances can change their state under extreme pressure; for example, water and many other liquids solidify. In rarer instances the reverse can happen. Solids can transform from one crystal structure to another in order to optimize the packing of atoms. Transparent salts turn into black metals. Magnetic materials such as iron lose their magnetism. The higher the pressure, the longer the list of surprises. Put another way, under extreme pressure chemical bonding is profoundly changed, such that a completely new periodic table emerges: Potassium becomes a transition metal and oxygen a superconductor.

The work of Mao and Hemley is part of a resurgence of interest in high-pressure experiments. They are members of a second generation of investigators in a field that was believed to have reached its maturity more than 50 years ago with the Nobel prize awarded in 1946 to pioneer Percy W. Bridgman. Hundreds of substances had been compressed in Bridgman's ingenious apparatus, up to pressures exceeding 100,000 atmospheres. Bridgman's successors have achieved new results both with static-compression methods, such as the diamond-anvil cell, and also with the refinement of dynamic-compression methods based on shock waves. As soon as new record pressures are announced, new and surprising phenomena are discovered. In 1976, Mao and Bell broke the one-million-atmosphere barrier. Breaking the barrier was not merely a symbolic event. It meant they were able to reproduce the pressures at the bottom of the Earth's mantle and deep inside the giant planets.

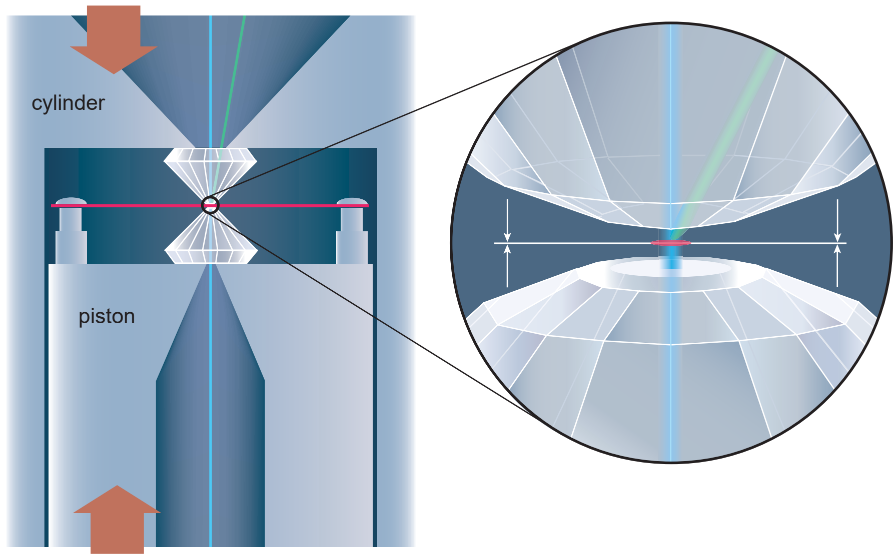


Figure 3. Diamond-anvil cells squeeze tiny samples of matter between a pair of gems at pressures close to those known to exist in planetary interiors—millions of times Earth's atmospheric pressure. In the device, which is about the size of a standard box of tissues, pistons and screws apply pressures that are sometimes great enough to fracture diamonds, the hardest known material. A laser or x-ray beam *(blue)* is scattered *(green)* to read detailed information from a tiny sample, detecting alterations in the material that are often dramatic.

Metallic Hydrogen

Back in 1935, Eugene Wigner, one of the founding fathers of quantum mechanics and at the time a professor at Princeton University, suggested that hydrogen, an inert molecular gas at ambient conditions, could turn into a metallic solid, similar to lithium or sodium, at sufficiently high pressure. Wigner's proposal implied a remarkable complexity for "element one," the simplest chemical entity, one electron bound to one proton.

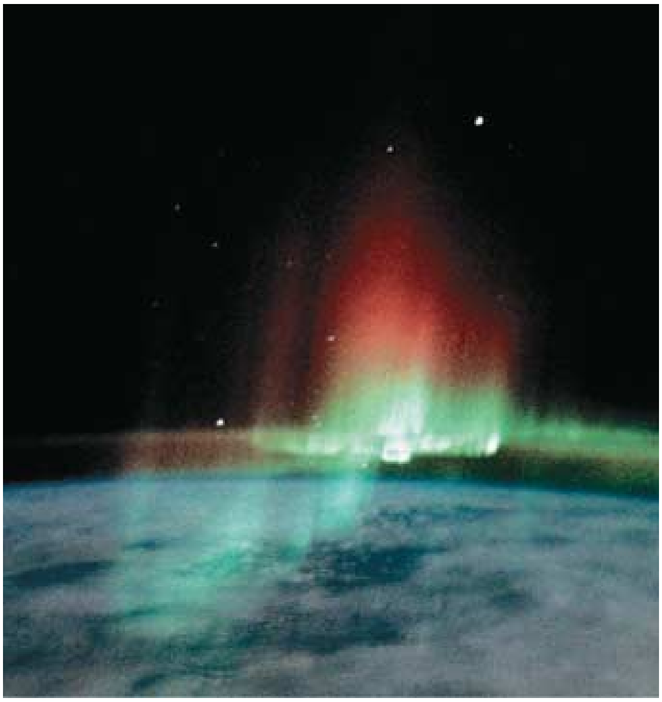
Because hydrogen is known to make up about 90 percent of the volume of Jupiter and Saturn, the appearance of a metallic state of hydrogen at high pressure could seriously alter our understanding of planetary interiors. Planetary and stellar magnetic fields are generated through a dynamo-like mechanism by electrical currents in the metallic regions of their interiors. Earth's magnetic field, for example, originates in the liquid metallic outer core. Jupiter's magnetic field, first measured by Voyager spacecraft, is ten times stronger than Earth's, and its pattern is considerably more complex. Part of this complexity could be accounted for if the source of the field lay much farther from the center, in relative terms, than does Earth's. Wigner's prediction of metallic hydrogen was based on a simplified analysis of the electronic ground state, but the pressure he calculated for the transition to the metallic state, about 250,000 atmospheres, corresponded to a depth of less than one-twentieth of the planetary radius of Jupiter. In other words, most of the solar system's largest gas giant had to be in a metallic state—although the metallic hydrogen would have to be a fluid rather than a solid to provide dynamo action.

Mao and Bell's achievements with the diamond-anvil cell immediately prompted high-pressure scientists to test Wigner's prediction and search for the metallic state of hydrogen. Unfortunately, a quarter-century later, and nearly 70 years after Wigner's proposal, no research group has been able to show conclusively that they have managed to turn hydrogen into a metallic solid under static compression in the laboratory, despite tremendous effort.

It turns out that Wigner's proposal, though probably correct at much higher pressure, was not entirely correct in detailing how and when metallization takes place. The emerging explanation lies in a subtle interplay between chemistry and physics.

In the periodic table of the elements, hydrogen is traditionally placed in the upper left corner, right above lithium and sodium. Column I of the table is where Dmitri Mendeleev, its originator, placed the alkali atoms—atoms with a single valence electron. The atomic state of hydrogen certainly meets this criterion. However, adding an electron to a hydrogen atom creates a rather stable ion, a criterion Mendeleev used to place atoms such as iodine on the opposite side of the periodic table in Column XVII.

Wigner's proposal relied heavily on this chemical ambiguity. At low density, the diatomic state of hydrogen (H_2), in which each hydrogen atom exhibits the behavior of a Column XVII element, is clearly preferred. But at sufficiently high compression hydrogen jumps across the table to Column I, where Mendeleev placed it. Unfortunately, a careful determination of the pressure at which this transition happens requires solving the quantum mechanics of the electrons and comparing their energy in the two states—the insulating diatomic state and the metallic monoatomic state. The basic equations of quantum mechanics had just been laid out in 1935 and already solved ex-



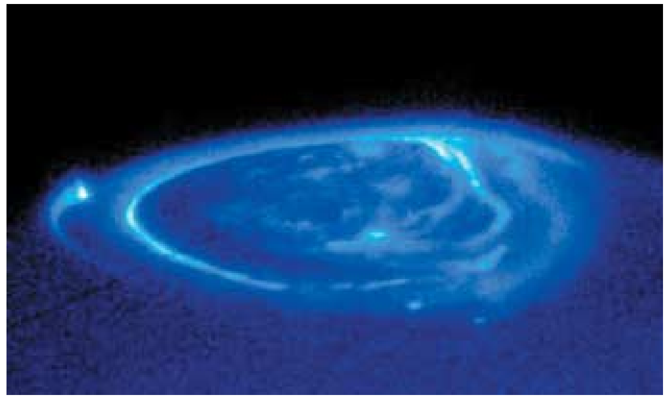


Figure 4. Planetary aurorae, the light displays produced by collisions between charged electrons in the solar wind and the atmosphere, illuminate the lines of magnetic force generated by planetary cores. Earth has aurorae near the South and North Poles; the latter, the aurora borealis, is seen at left in a Space Shuttle image. Jupiter's aurorae, one of which is shown in the Hubble Space Telescope image above, are more elaborate, revealing the stronger and more powerful magnetosphere surrounding Jupiter. The magnetosphere is part of the evidence indicating that much of Jupiter's interior must consist of hydrogen in a fluid metallic state. (Space Shuttle image courtesy of NASA; Hubble Space Telescope image courtesy of NASA/ESA and John Clarke.)

actly for a number of extremely simple cases, including the hydrogen atom itself. But solving such equations for a more complex case such as high-pressure solid, metallic hydrogen required huge approximations. Wigner ended up greatly underestimating the transition pressure.

Today, refinements in theory and extrapolations from experiments yield estimates indicating that hydrogen metallizes at pressures exceeding 4 million atmospheres—barely within the range of diamond-anvil cells. Moreover, it is currently believed that hydrogen metallization may be a matter more complex than a simple jump across the periodic table. Recent experiments have shown that iodine turns metallic while in the diatomic state (I_2) and becomes a monatomic alkali-like solid only at higher pressures. In other words, the route to metallic hydrogen might not be straightforward but may involve a sequence of transitions yet to be uncovered.

A Shocking Solution

The fact that hydrogen is reluctant to metallize on compression raised questions about our view of Jupiter. Is metallic hydrogen not so ubiquitous in Jupiter after all, but rather found limited to those areas close to the planetary core where the pressure is highest?

A close look at Figure 5 suggests a possible answer. The interiors of giant planets are in fact subjected to extreme pressures and extreme temperatures at the same time. Perhaps, theorists conjectured, temperature could play an unexpected role in metallization. Unfortunately this was a conjecture that could not be tested by studies using diamond-anvil cells. Heating materials inside a diamond-anvil cell is difficult, particularly in the case of hydrogen. Hot hydrogen tends to react with the gasket that holds it between the anvil tips as well as with the diamonds themselves. As a result, the highest temperature that has been reached in a diamond cell containing hydrogen is still below 850 kelvins—although, as we note below, important studies of combined pressure-temperature effects in other elements have been accomplished with diamond cells. (A kelvin, a degree on the Kelvin temperature scale, is equal to a centigrade degree, but the scale begins at absolute zero, or –273.15 degrees Celsius.)

Compressing hydrogen with shock waves seemed a more promising approach to the temperature question. Indeed, shock-wave experiments suffer from the opposite problem. Pressures in the million-atmosphere range can only be reached with an intense shock wave, of the sort generated when a metal projectile or an extremely intense pulse of laser light smashes into a sample. But the more intense the shock, the higher the final temperature of the sample. When directly shocked to a million atmospheres, hydrogen heats up to temperatures in excess of 20,000 kelvins, far above the range of temperatures estimated for the corresponding depths of the planetary interior.

But in 1995, Bill Nellis, Sam Weir, Arthur Mitchell and their coworkers at Lawrence Livermore National Laboratory managed to design and operate a shock-wave apparatus that was improved with a couple of old tricks of the trade. First they cooled down the pre-shocked sample so as to increase its density and bring it closer to the target value. Second, they designed the apparatus in such a way that the shock wave would reverberate between the projectile and the chamber walls.

Calculations predicted that much higher pressures could be reached with a reverberating shock, and without so large a temperature increase. At variance with diamond-anvil experiments, where the sample can be kept in a compressed state for an unlimited time, measurements in a shock-wave experiment must be carried out rather quickly. In less than one microsecond the whole sample assembly blows up, incinerated by the blast. But Nellis's team finally managed to measure the electrical conductivity of hydrogen up to 1.8 million atmospheres and 2,900 kelvins, very close to jovian-core temperature and pressure conditions, and found that hydrogen turns metallic at 1.4 million atmospheres and 2,600 kelvins, less than half the pressure plausibly required to metallize it at room temperature. Neither Wigner nor the diamond-cell scientists engaged in the search for metallic hydrogen could have anticipated that the effect of high temperature would be so dramatic. But the final picture of Jupiter that emerged from the shock experiments was quite neat. The measured conductivity and the new estimate of the transition pressure to metallization were consistent with the strength and pattern of Jupiter's surface magnetic field. Every brick of the model was now falling into its proper place, from the microscopic scale of the shock-wave experiment, to the planetary scale of the magnetic field generation.

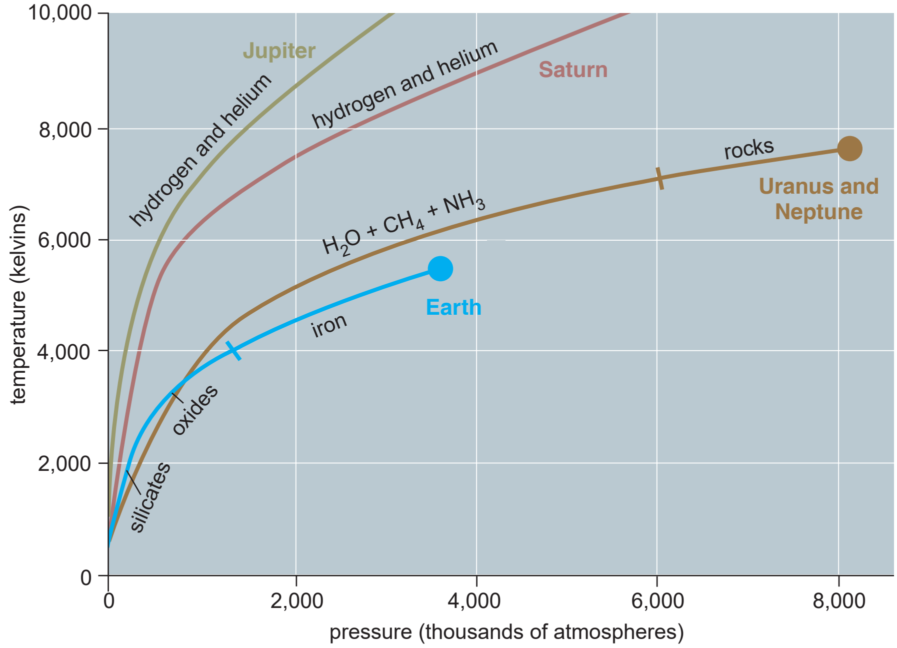


Figure 5. Experiments have begun to shed light on how temperature and pressure may interact to create unusual states of matter in planetary interiors. In the gas giants, temperature is thought to play a significant role. Shock-wave experiments suggest that high temperature causes hydrogen to metallize at a fraction of the pressure required to cause this transition at room temperature, supporting predictions that much of Jupiter may consist of metallic hydrogen. Temperatures are not believed to reach similar heights in the interiors of Uranus and Neptune, which can be more closely modeled with diamond-anvil experiments and computer simulations. Likewise the pressure and temperature conditions at the center of Earth are within range of diamond-anvil experiments incorporating laser heating of iron, but technical issues prompt continuing debate about interpretation of the experimental results.

Diamonds in the Sky

Neptune and Uranus lie near the borders of the solar system, a few billion kilometers from the Sun. It is not surprising, then, that the first serious attempts to model the interiors of these planets began only on the occasion of the Voyager II fly-by, less than 20 years ago. Yet, based on their density and distance from the Sun, scientists have long speculated that the interiors of Neptune and Uranus must be compositionally more rich than those of Jupiter and Saturn—with water, ammonia and methane, the so-called planetary ices, contributing about 80 percent of the mass of each planet. Indeed, spectroscopic studies conclusively reveal the presence of these molecules in the outer atmospheres of these planets, as well as in the atmospheres of small stars known as "brown dwarfs."

To be fair, little was known about the actual state of these molecular ices at deep-planetary conditions before scientists started to reproduce Neptune's pressures and temperatures in the laboratory. So it was startling news indeed when Marvin Ross, analyzing fresh shock-wave data on methane taken by his colleagues at Livermore, announced in 1981 that a giant mine of diamonds could hide in the core of Neptune. Methane is composed of one carbon and four hydrogen atoms (CH_4), but extreme compression, Ross argued, was causing the molecule to completely dissociate, and its carbon atoms would re-aggregate into their most stable form at those conditions—diamond. Al-

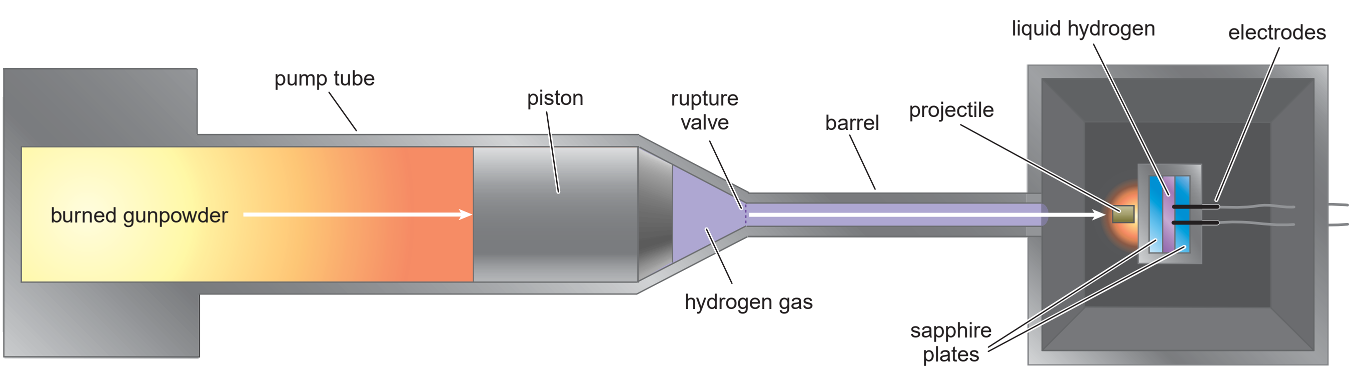


Figure 6. Shock-wave experiments compress matter to high pressures and high temperatures, using a laser light pulse or projectile to smash a sample. In the mid-1990s, a team at Lawrence Livermore National Laboratory built an apparatus capable of shocking a hydrogen sample to a pressure of 1.8 million atmospheres and a temperature of 2,900 kelvins, close to the conditions in the jovian core. Their experiments showed that high temperature dramatically reduced the pressure needed to cause hydrogen to metallize, yielding support for the notion that much of Jupiter's interior consists of metallic hydrogen. In the shock-wave "gun," a liquid-hydrogen sample is cooled and placed in a holder. Hot gases from a gunpowder explosion propel a piston that compresses hydrogen gas in a piston tube; the gas rushes into the barrel of the gun to propel a projectile toward the sample. The intense shock of the impact subjects the sample, for an instant, to planetary-interior conditions.

though there was no doubt that this had to be the fate of methane in the deepest regions of Neptune, the question remained as to whether methane had to be completely erased from the list of components of the planetary interior. A hint of the answer to this question came only in 1996, but it came neither from shock waves nor from diamond-anvil experiments. In fact it came not from experiment but from a radically different way of simulating planetary interiors.

Extracting information about the large-scale composition of a planet from a shock lasting less than a millionth of a second or from a squeezed sample weighing a millionth of a gram was a giant leap for planetary science and a fascinating example of scientific endeavor. But because the laws of nature must hold down to the atomic scale, there is no reason why such an experiment cannot be miniaturized even further, to the point that the sample consists of just a few molecules. This is the scale where available theoretical methodologies and current computing facilities allow physicists and chemists to solve the basic equations that govern the behavior of electrons and atoms in matter and provide a detailed picture of how atoms bounce into one another, vibrate and get squeezed under the combined action of pressure and temperature.

The idea of simulating the behavior of matter at the atomic scale is actually as old as the computer itself. Enrico Fermi, Stanislaw Ulam and John Pasta were probably the first to recognize, in 1955, the potential benefit of using computers to solve Newton's equations of motion. They solved the real-time dynamics of a collection of interacting point masses coupled with springs—a highly idealized system indeed. But methodological developments, theoretical advances in our understanding of how atoms interact (through quantum mechanics) and, not least, breathtaking increases in the speed of computers have brought us to a point where the idea of simulating a bunch of atoms from the bottom up—by solving exactly the laws of quantum and classical mechanics—has become as feasible as squeezing the real material in a diamond-anvil cell or in a shock-wave apparatus.

So it was that in 1996, one of us (Scandolo), with colleagues in Trieste, Italy, set about to simulate on the computer the fate of methane at the conditions of pressure and temperature of the interior of Neptune.

Virtual Neptune

In concept, a simulation of the behavior of methane at planetary conditions does not differ tremendously from what Fermi, Pasta and Ulam had done four decades earlier. The Trieste group took a bunch of molecules—16 was the maximum number we could afford with the supercomputers available at the time—put them in a simulation cell and let the positions of the atoms evolve according to Newton's equations—that is, with an acceleration equal to the force divided by the atomic mass.

Newton's equations are solved in this case by dividing time into very short intervals, each less than a femtosecond (10^{-15} second) long, calculating forces at every time step and updating the atomic positions accordingly. One picosecond (10^{-12} second) of dynamics requires repeating this operation more than a thousand times. We needed a supercomputer instead of the rudimentary punch-card machines available to our 1950s predecessors because the force ex-

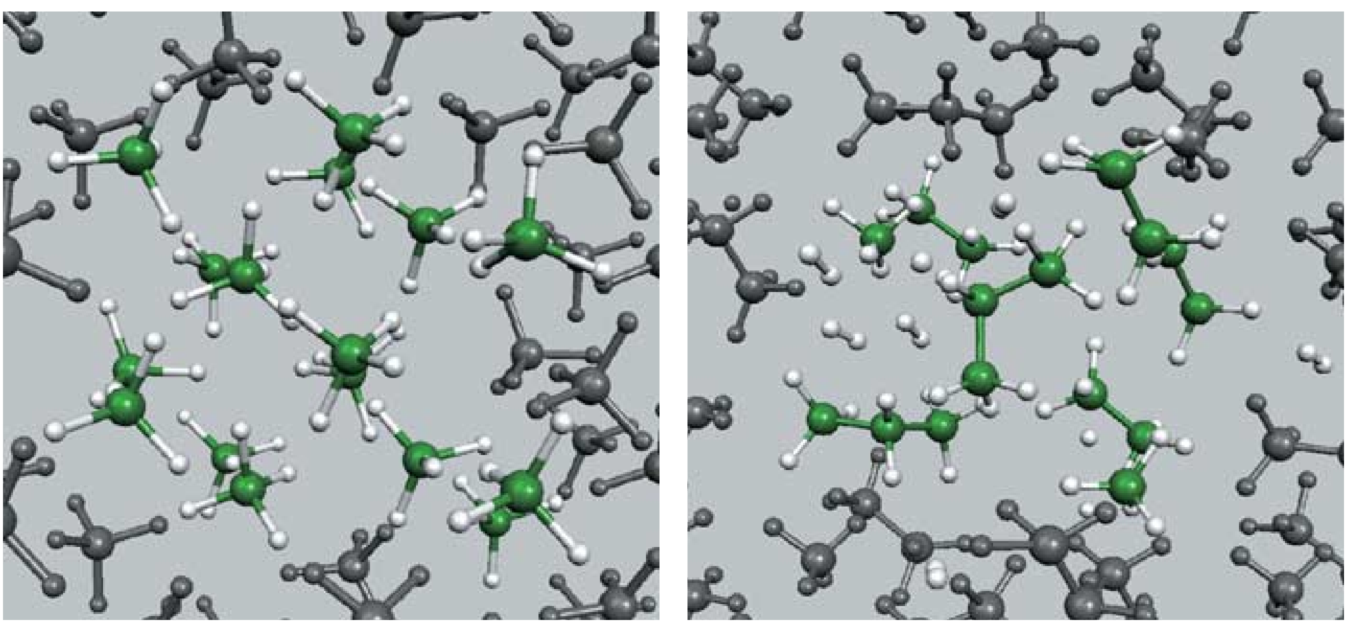


Figure 7. Could the center of Neptune be filled with diamonds? Shock-wave experiments with methane, which has the chemical formula CH_4, suggested that extreme pressures could cause methane to dissociate—separating its carbon and hydrogen atoms. Under such conditions the carbon atoms would be expected to aggregate to form diamond, carbon's most stable form. Support for this prediction came from computer simulations. Author Scandolo and colleagues at the Abdus Salam International Centre for Theoretical Physics in Trieste, Italy, simulated the dynamics of 16 methane molecules and found that diamonds indeed formed at the high pressure–high temperature conditions of Neptune. At intermediate pressures, however, methane dissociated partially and formed hydrocarbon chains. Snapshots of this simulation are shown above. At left the original 16 methane molecules (one green carbon atom seen attached to four white hydrogens) are seen at conditions of relatively low temperature and pressure. After one picosecond at 4,000 kelvins temperature and 100 gigapascals pressure (about 1 million atmospheres), the molecules have dissociated and recombined, forming two methane, four ethane (C_2H_6) and two propane (C_3H_8) molecules, with extra hydrogens left mainly as diatomic molecules. (Reprinted from Ancilotto *et al.* 1997, by permission of the American Association for the Advancement of Science.)

erted by one atom on another atom cannot simply be modeled by a spring, as Fermi, Pasta and Ulam had postulated. Interactions between atoms are mediated by the presence of their electronic clouds. Electrons rearrange instantaneously with every change in the positions of the atoms and, depending on the external conditions applied to the system, can either hold atoms together as a kind of glue (a chemical bond) or cause their separation, as in the case of molecular dissociation.

Following the rearrangements of the electronic clouds, and thus calculating the forces acting on atoms, is an extremely difficult task that involves solving the quantum mechanics of hundreds of electrons simultaneously and repeating the operation as many times as the atomic dynamics require. It was no surprise, then, that two weeks of supercomputer time were needed to simulate just five picoseconds of the "real" dynamics of 16 methane molecules. Fortunately, chemical reactions such as dissociation take place very rapidly, typically on the femtosecond timescale, so we would not miss them if they happened in our simulated environment.

Mining this small virtual world, we found Ross's diamonds. The results of the simulations confirmed Ross's proposal that diamonds form under the conditions found in the deepest regions of Neptune. But the calculations unexpectedly yielded a different picture at intermediate pressures, those corresponding to the bulk of the planet. Instead of breaking down completely into its atomic constituents, methane in the simulation dissociated only partially and ended up forming hydrocarbon chains, chains of two to three carbon atoms surrounded by hydrogen atoms. The discovery added strength to Ross's idea that methane had to be eliminated from the list of "ices," and it implied that Neptune's deep chemistry had to be more complex than previously thought. In particular, the production of hydrocarbons in the planetary interiors could account for the observed anomalous abundance of some of these substances in the atmosphere of the planet, where they might be brought up from the deep interior by convective currents.

Direct experimental confirmation of both hydrocarbon and diamond formation from methane at planetary conditions came only three years later, in 1999, from a diamond-anvil experiment carried out in Berkeley, California, by one of us (Jeanloz), with Robin Benedetti and other coworkers. Real diamonds popped out, floating in a bath of fluid hydrocarbons, when a methane sample was heated above 2,500 kelvins and compressed above 200,000 atmospheres in the diamond cell.

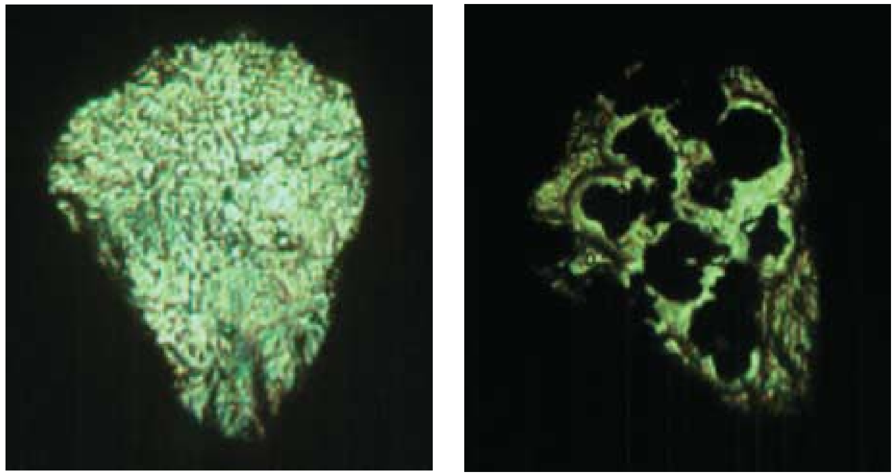


Figure 8. Experimental confirmation that hydrocarbons and diamonds could both form from methane at planetary conditions came from a diamond-anvil experiment carried out at the University of California, Berkeley, by author Jeanloz and coworkers. A methane sample is shown here in photomicrographs taken before *(left)* and after squeezing and laser heating in the diamond cell. In measurements of the infrared absorption spectra taken afterward, the signature of methane had faded, replaced by absorbance bands characteristic of doubly and triply bonded carbon in hydrocarbons. At the center of the laser beam, where heating was most intense, there was evidence of diamonds. (Image from Benedetti *et al.* 1999, reprinted by permission of the American Association for the Advancement of Science.)

This figure was even lower than the required pressure for methane dissociation predicted by the computer simulations, which implies that perhaps no methane at all can be found deep inside Neptune. The findings have additional implications. The separation of methane into rising hydrogen and sinking diamond likely releases gravitational energy to drive the convective motions of the planet's fluid interior. The amount of this energy appears to be large, comparable to the excess heat—over and above the heat received from the Sun—that infrared emissions indicate is released from Neptune's interior.

Earth's Hottest Dispute

Our virtual journey to planetary interiors finally brings us home to Earth, the spaceship on which we reside. Although Earth is the most studied of all planets, its interior is still profoundly mysterious. It is also remarkably inaccessible. Yet the interior holds key information about how our planet formed and evolved over geological time, motivating decades of high-pressure experiments.

One of the most immediate questions is: How hot is the deep interior? It is the heat of the Earth's mantle and core that causes geological activity, from volcanic eruptions and the movement of continents to earthquakes and the deposit of ore bodies. Much of that heat is left over from the formation of our planet, 4.5 billion years ago; additional heat comes from the decay of naturally occurring radioactive isotopes of elements such as potassium, thorium and uranium. It continues to drive the geological evolution of our spaceship.

The most direct way to answer this question is to determine the melting temperature of the material in the Earth's core at high pressures. Measurements of seismic waves passing through the interior show that the outer core is liquid (with a viscosity thought to be comparable to that of the oceans), whereas the increase in pressure with depth causes the inner core to be solidified. Therefore, the interface between the inner and outer core must be at the freezing (or, if you prefer, melting) temperature of the core material at that depth. Because of the fluid nature of the deep interior, seismological measurements and the equations of fluid mechanics can be used to calculate the pressure at this boundary: 3.25 million atmospheres.

If we think of Earth as a huge press capable of showing us (were we able to insert a thermometer) the freezing temperature of the core alloy at high pressures, we can imagine building a miniature version of this press in the laboratory to measure the melting and freezing temperatures of appropriate alloys at pressures in the range of 3 million to 4 million atmospheres. Were this possible, we could determine the temperature at the inner core–outer core boundary and, by a modest extrapolation, right to the center of the planet.

Michael Brown, then a graduate student at the University of Minnesota, started going to Los Alamos National Laboratory in the late 1970s to work with Robert McQueen, a leader in shock-wave experiments. Using methods that had been pioneered at the laboratory, Brown and McQueen showed that iron melts when shock-compressed to pressures of about 2.5 million atmospheres. They discovered that although the speed of sound in iron increases as the sample is shocked to higher pressures, the sound velocity drops at 2.5 million atmospheres in exactly the manner that would be expected for melting, thereafter increasing as the (molten) iron is shocked to higher pressures.

Published in 1982, these findings nicely paralleled those of Danish seismologist Inge Lehman, who discovered the inner core in 1936 by determining that the velocity of seismic waves abruptly increases at a depth that we now identify as the interface between the solid and liquid regions of the core.

Unfortunately, temperature could not be readily measured in the shock experiments at Los Alamos. Still, the discovery of the melting transition at high pressures was a major advance that motivated other investigators. Happily, unlike hydrogen, iron can be heated in a diamond-anvil cell using a laser beam. Within a few years, Quentin Williams and one of us (Jeanloz) were measuring the temperature of laser-heated iron at high pressures at the University of California, Berkeley. By measuring the spectrum of the light emitted from the hot sample, we could gauge temperature by the same methods astronomers use to determine the surface temperatures of stars. The results were surprising: Instead of melting at about 3,000 kelvins, as expected, it seemed that iron required temperatures closer to 4,000 kelvins in order to melt at 1 million atmospheres of pressure.

At the same time, Thomas Ahrens, Jay Bass and their associates at Caltech had managed to use the same method as the Berkeley group to measure the temperature of iron as it is shock-compressed to 3 million atmospheres. They again found a surprisingly high temperature at the shock-melting point of 2.5 million atmospheres, about 6,500 kelvins, in good accord with the laser-heated diamond-cell experiments.

But there were problems. First, a laser cannot uniformly heat a sample inside a diamond cell. Only the center of the hot spot at the focus of the laser beam reaches peak temperatures, and the temperature drops off to room temperature within less than 0.1 millimeter from the center. Typically the emitted light varies from "white-hot" at the center to "red-hot" and then dark (no visible emission) within a short distance across the sample. An experimenter trying to measure the spectrum from a small sample squeezed at high pressure between relatively thick diamonds faces a tough technical challenge. In addition, the most interesting part of the sample is at high temperatures. The sample glows so brightly that it becomes difficult to be sure whether or not it has melted.

Similarly, there were technical problems interpreting shock-wave results, because the hot iron sample has to be contained long enough at high pressures to be able to reliably determine the temperature. A window has to be put on the back side of the sample, altering both the pressure and temperature achieved during the shock loading. Moreover, the experiment is over so quickly that even a sample at the melting temperature may not have time to melt; to achieve melting the experimenter might need to overshoot the true melting temperature, thus obtaining reproducible measurements that are consistently too high.

The agreement between static and dynamic experiments suggested that these difficulties had been overcome. However, the surprisingly high temperatures motivated others to try re-

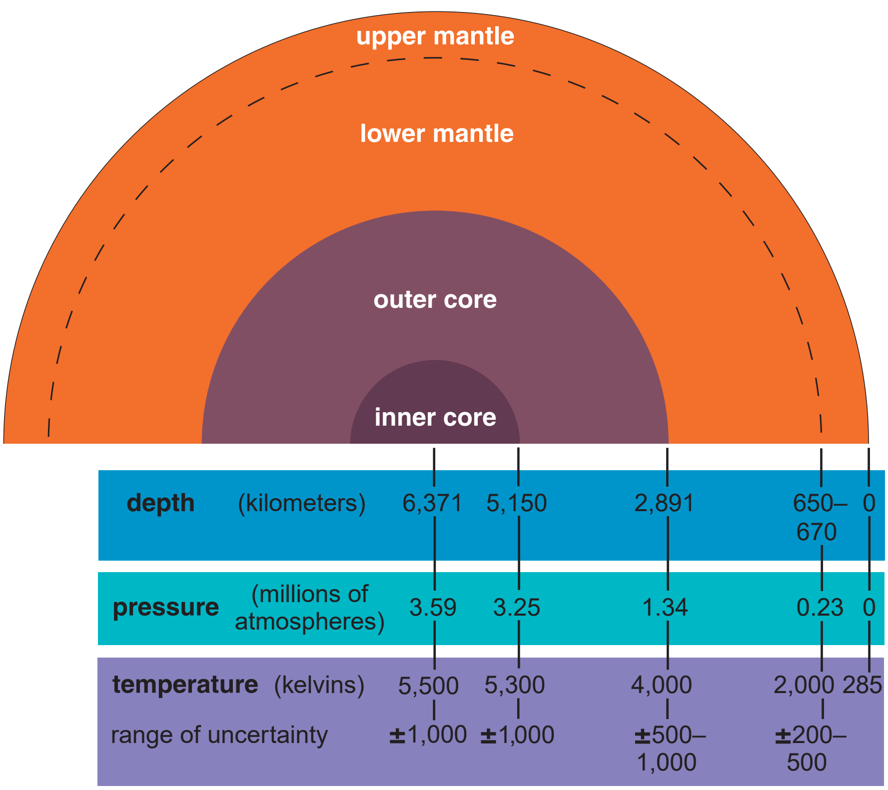


Figure 9. Debate continues over exactly what temperatures prevail at the boundary between Earth's liquid outer core and solid inner core. The pressures at boundaries between Earth's interior layers are known from seismological and other evidence, but attempts to simulate temperatures in the core have produced varying results. General agreement is emerging that the Earth's center may be as hot as the surface of the Sun—in the range of 5,000 to 6,000 kelvins. Here are shown temperature estimates derived from recent high-pressure experiments. More precise estimates of the temperature at the inner core–outer core boundary likely will require better modeling of the origin and evolution of the deep interior, which could reveal the role of alloying elements that could alter the melting point of iron.

producing the results. Problems quickly arose, and during the 1990s groups in Germany, Sweden and the U.S. reported a variety of melting temperatures as they varied experimental parameters. Controversy arose as there were indications for and then against (and corresponding claims in the scientific literature) a new crystalline form of iron having been discovered at high pressures. In order to make sense of this large and confusing array of new data, the various groups have been refining their methods and applying ever more sophisticated tools.

Our picture of the inner core–outer core boundary will no doubt evolve as different methods are used to check these findings, and as refinements in laboratory techniques result in smaller experimental uncertainties. Good calibration standards have yet to be developed for measuring temperatures (let alone melting temperatures) in the 3,000- to 5,000-kelvin range. But this also leaves us with the question: How good is "good enough"?

We may be nearly there. The fact is that the Earth's core is not pure iron but contains about 10 percent (by weight) of other constituents. If you compare the density of the outer core that is derived from seismological data with that of pure iron shocked to comparable pressures and temperatures, the core's density turns out to be about 10 percent lower. Even when the melting temperature of pure iron is accurately known at 2 million to 4 million atmospheres of pressure, we will still have to make a correction for the effect of contaminants. Alloying often decreases the freezing temperature of a material; this is why ice can be melted by putting salt on top of it. The actual freezing temperature at the inner–outer core boundary may therefore be 1,000 kelvins or so lower than that of pure iron.

Yet the exact makeup of the core alloy is impossible to know. The current composition of the core is the result of the processes by which it first formed and subsequently evolved over geological time. There are many competing ideas: Carbon and sulfur, oxygen and even hydrogen have been proposed as candidates for the primary alloying component. High-pressure melting studies of such alloys are ongoing. It is already clear that the addition of hydrogen or sulfur may significantly lower the melting temperature of iron, but this is not the case for other alloying components. A good model for the origin and evolution of the Earth's deep interior will be required before we can determine the compositions relevant for experimental study and ultimately make a good estimate of the temperature at depth.

The current uncertainty over the core's composition thus parallels the uncertainty resulting from the various experimental results, which, although somewhat scattered, are in general agreement. An interplay between the two—where a refined understanding of the evolution and composition of the core drives new experiments to determine the behavior of alloys at high pressures—seems most likely to answer large questions about the center of Earth.

(Still) Having a Heat Wave

Broadly speaking, however, these experiments have truly rewritten the texts about the Earth's interior. Before the shock-wave and diamond-cell experiments, estimates of core temperatures were little more than educated guesses. From values of 3,500 to 4,300 kelvins, estimates of the central temperature have nearly doubled to 5,500–6,000 kelvins. To be sure, the uncertainty on this estimate is as big as ever—about 1,000 kelvins in either direction—but the effects of alloying and experimental uncertainties are being factored into estimates that are now based on measurement.

It turns out that the temperature at the center of our planet is likely to be comparable to that of the glowing-hot surface of the Sun. How did our planet get so hot in the first place? How has it managed to retain so much heat? Our rocky planet's mantle overturns itself over geological time, like a thick gravy heating in a pan on the stove. But the flame is on low—it is thought that there is only relatively modest heating from natural radioactivity at great depth—so how can it be that our planet has not cooled itself off by now, and that it remains so geologically vigorous? The greatest surprises may lie not within the mysterious planets in the sky, but within the roiling, boiling one beneath our feet.

Bibliography

Ancilotto, F., G. L. Chiarotti, S. Scandolo and E. Tosatti. 1997. Dissociation of methane at extreme (planetary) pressure and temperature. *Science* 275:1288–1290.

Benedetti, L. R., J. H. Nguyen, W. A. Caldwell, H. Liu, M. Kruger and R. Jeanloz. 1999. Dissociation of CH_4 at high pressure and temperature: Diamond formation in giant planet interiors? *Science* 286:100–102.

Boehler, R. 1993. Temperatures in the Earth's core from melting-point measurements of iron at high static pressures. *Nature* 363:534–536.

Brown, J. M., and R. G. McQueen. 1986. Phase transitions, Grüneisen parameters and elasticity for shocked iron between 77 GPa and 400 GPa. *Journal of Geophysical Research* 91:7485–7494.

Fermi, E., J. Pasta and S. Ulam. 1955. *Studies in Nonlinear Problems, I.* Los Alamos report LA 1940.

Hemley, R. J., and N. W. Ashcroft. 1998. The revealing role of pressure in the condensed matter sciences. *Physics Today* (August) 51(8):26–32.

Loubeyre, P. 1999. Une pluie de diamants à l'intérieur de Neptune? *La Recherche* 326:18–19.

Mao, H. K., and P. M. Bell. 1976. High-pressure physics: the 1-megabar mark on the ruby R_1 static pressure scale. *Science* 191:851–852.

Nellis, W. J. 2000. Making metallic hydrogen. *Scientific American* (May) 282(5):84–90.

Ross, M. 1981. The ice layer in Uranus and Neptune—Diamonds in the sky? *Nature* 292:435–426.

Saxena, S. K., G. Shen and P. Lazor. 1994. Temperature in the Earth's core based on melting and phase transformation experiments on iron. *Science* 264:405–407.

Scandolo, S., G. L. Chiarotti and E. Tosatti. 2000. Diamonds in the sky? *Physics World* (October) 13(10):31–36.

Shen, G., H. K. Mao, R. J. Hemley and M. L. Rivers. 1998. Melting and crystal structure of iron at high pressure. *Geophysical Research Letters* 25:373–376.

Stevenson, D. 2003. Mission to Earth's core—a modest proposal. *Nature* 423:239–240.

Weir, S. T., A. C. Mitchell and W. J. Nellis. 1996. Metallization of fluid molecular hydrogen at 140 GPa. *Physical Review Letters* 76:1860–1863.

Wigner, E., and H. B. Huntington. 1935. On the possibility of a metallic modification of hydrogen. *Journal of Chemical Physics* 3:764–770.

Williams, Q., R. Jeanloz, J. Bass, B. Svendson and T. J. Ahrens. 1987. Melting curve of iron to 250 GPa: A constraint on the temperature of the earth's center. *Science* 236:181–182.

Yoo, C. S., N. C. Holmes, M. Ross, D. Webb and C. Pike. 1993. Shock temperatures and melting of iron at Earth core conditions. *Physical Review Letters* 70:3931–3934.

For relevant Web links, consult this issue of *American Scientist Online:*

http://www.americanscientist.org/IssueTOC/issue/406

On Neptune, It's Raining Diamonds

Deep within Neptune and Uranus, it rains diamonds—or so astronomers and physicists have suspected for nearly 40 years. The outer planets of our Solar System are hard to study, however. Only a single space mission, Voyager 2, has flown by to reveal some of their secrets, so diamond rain has remained only a hypothesis.

Beyond the lingering mystery of the diamond rain, there's a big loss in our failure to study Uranus and Neptune inside and out. It limits our understanding of the Solar System and the galaxy, because planets of this size have turned out to be extremely common in the Milky Way. The number of planets similar in size to Uranus and Neptune that have been found in the galaxy is roughly nine times greater than the number of much larger planets similar in size to Jupiter and Saturn. The outermost planets also seem to bear scars that could tell us a lot about the formation of our own Solar System (*see page 280*). So there's a growing sense of urgency to explore Neptune and Uranus—both to better understand where and how planetary systems form and also to refine our ideas about where to look for planets that can sustain life.

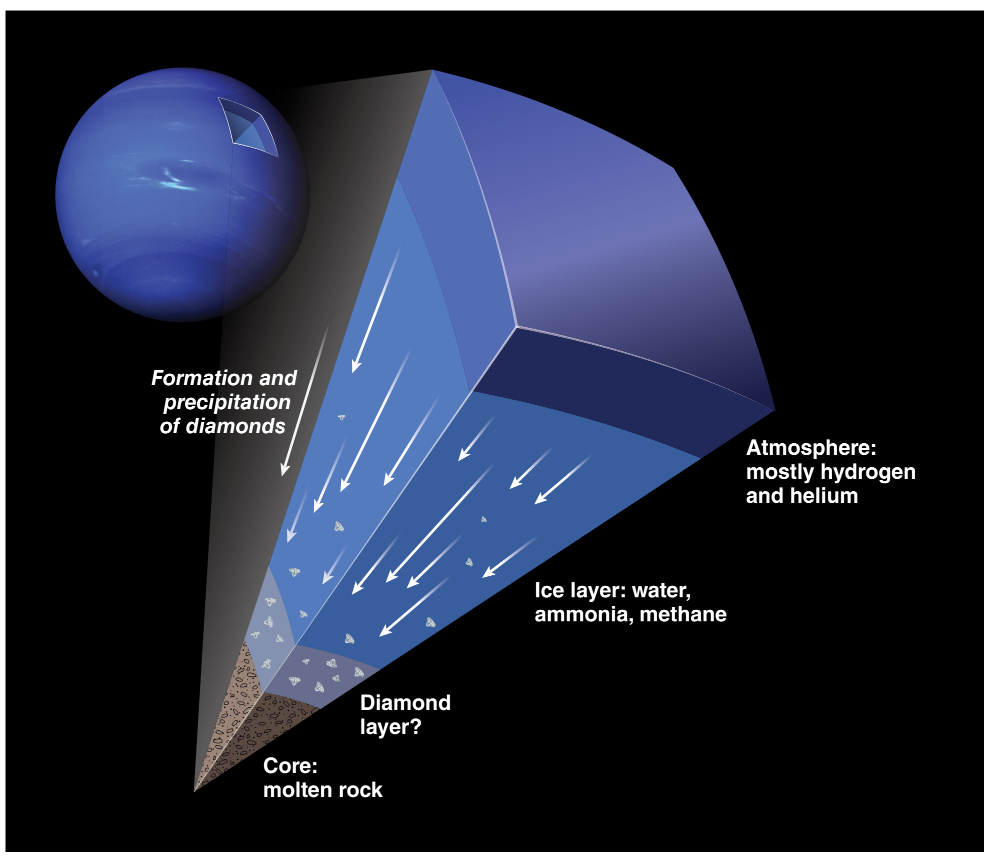
Although we've been limited by spacecraft and ground-based telescopes regarding how much we can learn about the exteriors of Uranus and Neptune, advances in laboratory simulations are enabling remarkable new insights about what's happening in their interiors, including what gives rise to diamond rain. Discoveries such as these reveal the complexity of the chemical processes involved in the evolution of these planets. Our simulations give clues to the internal nature of worlds far beyond the Solar System, even worlds that we may never see directly from the outside.

The Ice Giants

Neptune and Uranus are called the "ice giants" of our Solar System because their outer two layers consist of compounds that include hydrogen and helium. In astronomy slang, *ice* refers to all compounds of light elements that contain hydrogen, so the planets' water (H_2O), ammonia (NH_3) and methane (CH_4) make them "icy." The beautiful bluish hue of both planets is the result of methane traces in their atmospheres.

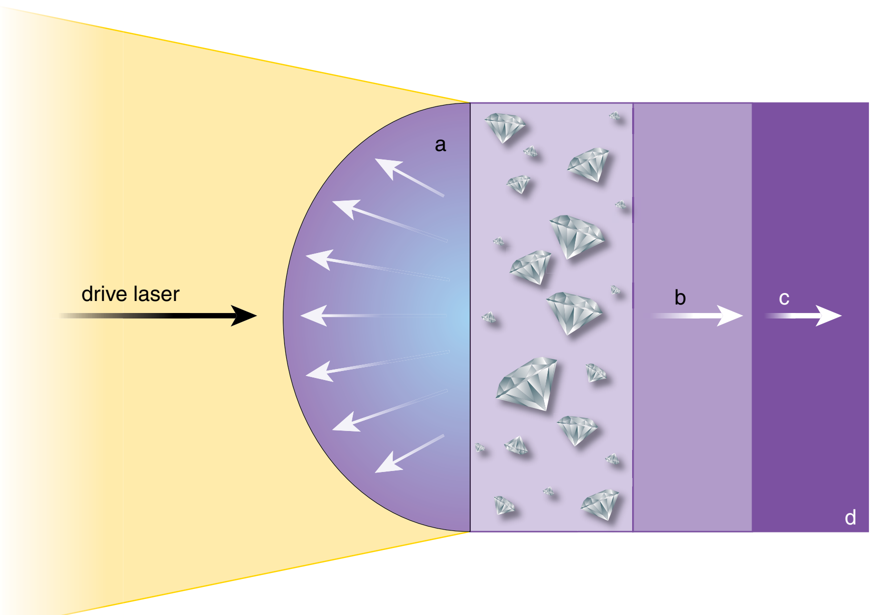
However, it is the "ice" in the deep middle layers that really shapes their properties. On Neptune, for example, beneath a hydrogen-helium atmosphere that is 3,000 kilometers thick lies an ice layer that is 17,500 kilometers thick. Simulations suggest that gravity compresses the "ices" in this middle layer to high densities, and the internal heat raises the internal temperatures to several thousand kelvins. Despite the high temperature, pressures more than one million times greater than the atmospheric pressure on Earth compress the so-called ices into a hot, dense fluid.

Under such heat and pressures, ammonia and methane are chemically reactive. Scientists have modeled exotic processes—including diamond formation—taking place between the compounds deep within the ice layers. Marvin Ross of Lawrence Livermore National Laboratory first introduced the diamond-rain idea in a 1981 article in *Nature* titled, "The Ice Layer of Uranus and Neptune—Diamonds in the Sky?" He suggested that the carbon and hydrogen atoms of hydrocarbons



Jason McAlexander

This cutaway of Neptune shows the different layers of the planet, including one where diamonds may form. Models suggest the diamonds collect in a layer just above the planet's core.



Researchers at Stanford's Linac Coherent Light Source work on the Matter in Extreme Conditions experiment (*top*). In the experiment (illustration, *bottom*), a fast-expanding, hot plasma (*a*) is created when an intense laser hits a polystyrene sample (*d*). Two subsequent compression waves travel through the sample. Diamonds form in regions where two waves have moved through the sample (*b*, *c*). The conditions in that region are comparable to the interior of the ice giant planets.

U.S. Department of Energy; Jason McAlexander

such as methane separate at the high pressures and high temperatures inside the ice giant planets. Clusters of isolated carbon atoms would then be squeezed into a diamond structure, which is the most stable form of carbon under such conditions.

Diamond is denser than the methane, ammonia, and water left in the ice layer, so the carbon crystal would start to sink toward the planet's core. It would accumulate new layers as it falls when it touches other isolated carbon atoms or diamonds, allowing individual diamond blocks to reach a size meters in diameter. We think that, as a result, a thick layer of carbon surrounds the rocky cores of Uranus and Neptune. This carbon layer may consist of blocks of solid diamond—or, if the temperature is extremely high (as some planet models suggest), it might transform into liquid carbon, or a mix of solid carbon and liquid carbon.

If the layer is a mix of solid and liquid carbon, the solid carbon would be of lower density than the liquid, with the result that large "diamond bergs" would float on top of an ocean of liquid carbon. Each possible composition of the carbon layer—solid, liquid, or mixed—would affect the core of the planet differently. Solid diamond, for example, is electrically insulating and has a stiff crystal lattice, whereas liquid carbon is a metallic conductor and flexible. Determining the properties of the carbon layer could reveal whether or not Neptune and Uranus formed from a rocky protoplanet core billions of years ago.

Testing the Hypothesis

Although Ross's idea was certainly fascinating, it was mainly hypothetical at the time and needed to be verified by observations. It is impossible with any imaginable technology to design and build a probe that could penetrate deep into Neptune or Uranus and directly observe the formation of diamonds. Scientists instead have tried to recreate the extreme conditions of planetary interiors in their laboratories. Even this more limited goal is extremely challenging, since we need to reliably generate and measure pressures of several million atmospheres and temperatures of several thousand kelvins to simulate their effects on the elements found inside the ice giants. In essence, we need to build a piece of a planet in the lab.

Facilities around the world are tackling the problem by compressing a sample material, such as methane, between two diamond anvils with very small tips that press on the sample. The same effect of pressure enhancement can be seen on a different scale by placing something underneath the heel of a high-heeled shoe. Even though diamond anvils can generate pressures of several megabars (comparable to the pressure that would be produced by placing several thousand African elephants on top of that high-heeled shoe), the sample also needs to be heated by electrical currents or lasers in order to mimic hot planetary interiors. Using such a setup, some experiments have indeed formed diamond. However, in these setups the materials representing the planetary ice layers—methane, ammonia, or water—start to react with the diamond anvils and the gaskets. Those reactions can strongly alter and contaminate the results.

Another way to generate the extreme pressure and temperature conditions found inside the ice giant planets is to create *shock compression* using strong explosives, high-velocity gun projectile impacts, or pulsed high-energy lasers. Although this process both compresses and heats the sample at the same time, the samples remain in the interesting state for only a tiny fraction of a second. Particularly for the high-energy lasers, which can achieve gigabar pressures and temperatures of millions of kelvins (comparable to the temperature at the center of the Sun), the conditions usually last a few nanoseconds or less. That is a very limited time in which to obtain precise and direct measurements of structural changes of the sample.

This situation changed in 2009 with the completion of the world's first x-ray free-electron laser: the Linac Coherent Light Source at Stanford University. Combining this machine with a powerful pulsed-laser system allows us to study chemical reactions at conditions comparable to those in the deep interiors of giant planets in real time. Plastics, which are mainly made out of carbon and hydrogen, are useful substances to mimic the material mix in the ice layers of Neptune and Uranus.

In such experiments, the pulsed high-energy laser is focused on a spot 200 micrometers in diameter, which heats a thin surface layer of a plastic sample that is 80 micrometers thick. Its surface is instantaneously transformed into extremely hot plasma, with temperatures of several million kelvins. This plasma vapor expands rapidly. As a result, an extreme pressure force presses the remaining plastic material and drives strong compression waves into the sample. If tuned correctly, the experiment can precisely mimic the pressure and temperature conditions predicted inside ice giant planets.

These conditions only last for a billionth of a second, yet every single flash provides a precise snapshot of the chemical reactions inside the sample material. The experiments show that even on such extremely short timescales, chemical processes are fast enough to grow tiny diamonds from the carbon atoms inside the plastic samples. The formation rates observed in the lab suggest that within Uranus and Neptune, where there have been many millions of years to grow diamonds, meter-sized carbon crystals can form.

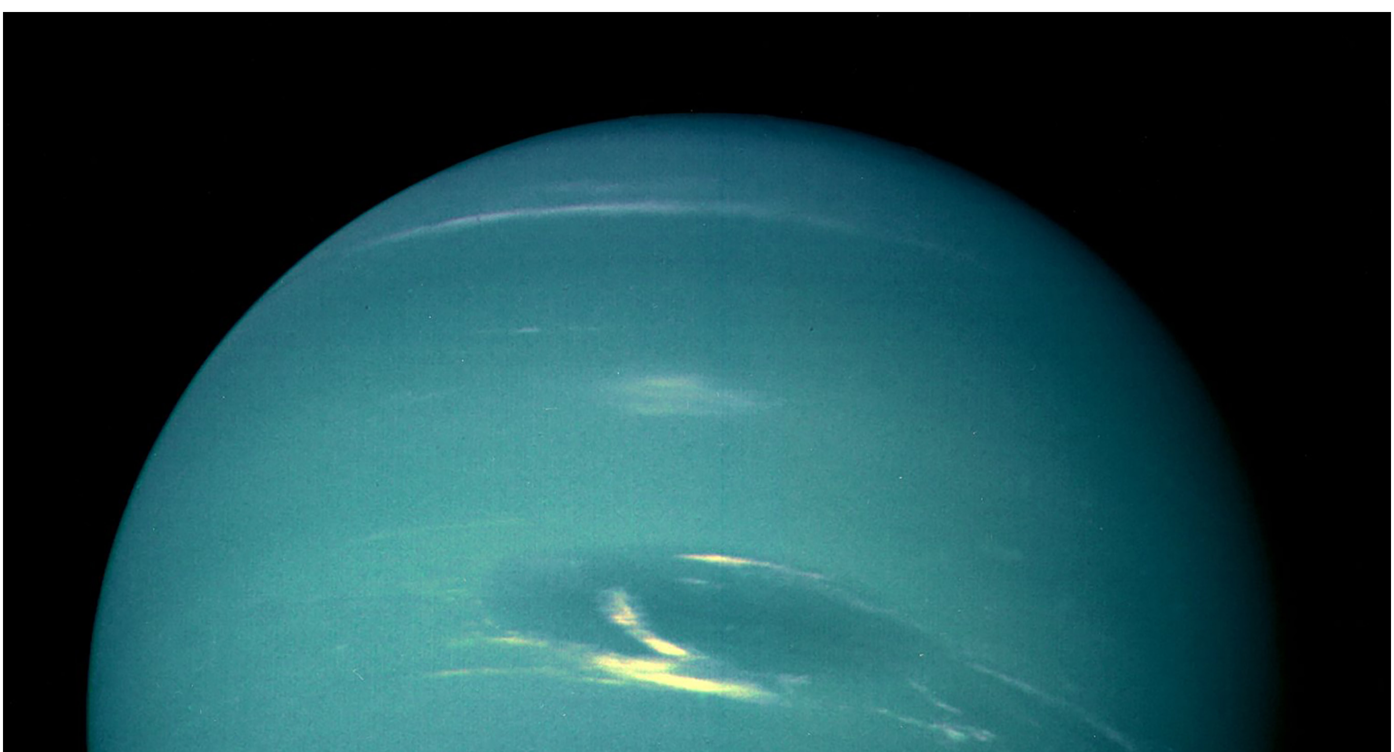
Diamond Influence

Understanding the inner processes of the ice giants gives clues to the features of these planets. For example, diamond precipitation releases gravitational energy, which is converted to heat by friction between the diamonds and the surrounding material as they descend. This effect could explain why Neptune is emitting more energy than it receives from the Sun. Such an internal energy source may help to account for the origin of the surprisingly violent storms that are observed on the planet's surface.

Diamond formation may also explain why the ice giant planets' magnetic fields are so exotic. Unlike Earth's magnetic field, the fields around Uranus and Neptune are not symmetrical, and they don't extend from each pole. These properties suggest that ice giant fields probably originate not in the core but in a thin, rather variable layer of conducting material, such as metallic hydrogen formed as a by-product of making diamonds. Other exotic processes inside the planets may also contribute to their magnetic fields. For instance, the formation of so-called *superionic structures* of water and ammonia, in which hydrogen ions can move freely through a crystal lattice of oxygen or nitrogen, could add to the conducting material of the magnetic fields.

We will continue to study these phenomena in the lab, but a new space probe mission to Neptune or Uranus (or both) could add a wealth of information about the planets' internal processes and about how such planets have formed in our Solar System and others. NASA is currently considering such a mission. In 2030, the planets of our Solar System will be favorably aligned for a spacecraft to launch and reach Uranus or Neptune by 2040. Another fortuitous alignment of the planets won't come for another two generations, so now is the time to start thinking about exploring the ice giants up close and learning more about the Solar System's intriguing diamond worlds.

Dominik Kraus is a Helmholtz Young Investigator Group Leader at Helmholtz-Zentrum Dresden-Rossendorf Institute of Radiation Physics and a Young Investigator at the Institute of Solid State Physics at Technische Universität Dresden, Germany. Email: d.kraus@hzdr.de

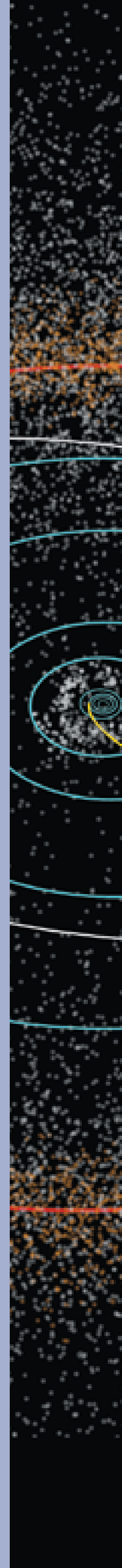


NASA

Journey to the Solar System's Third Zone

When New Horizons *reaches Pluto in July, it will close one era of space exploration and open an exciting new one.*

S. Alan Stern



This July, NASA's *New Horizons* spacecraft will complete a 9-year, 5-billion-kilometer journey from Earth to the frontier of the Solar System, where it will undertake the first close study of Pluto and its astonishingly diverse system of satellites. It will be a raw act of exploration unparalleled since NASA's *Voyager* missions to the giant planets in the late 1980s. Nothing quite like it has occurred in decades, and nothing like it is set to happen again in our lifetimes.

When most of us were taught basic astronomy in grade school, we learned that the Solar System consists of 4 inner rocky planets (the "terrestrials"), four outer giant, gaseous planets ("the Jovians"), and one small misfit: Pluto. But that was old-school science, limited by mid-20th century technologies that prevented us from seeing the cosmos as it truly is.

Beginning in the 1990s, planetary scientists—by then armed with large telescopes, high-sensitivity digital cameras, and fast computers—discovered that Pluto is no misfit at all. It is simply the brightest member of a vast population of objects orbiting beyond the Jovians: an entire third zone of the solar system. This region, first hypothesized in the 1940s by Gerard Kuiper, is now called the Kuiper Belt. It is littered with a diverse array of comets and small planets, of widely varying sizes. Pluto is both the largest (2,350 kilometers wide) of them and the first discovered, decades before the rest. The Kuiper Belt is, in turn, by far the largest zone of our planetary system.

New Horizons has flown for more than nine years to reach this distant shore. In the months around its closest approach on July 14 of this year, the probe will conduct a detailed survey of Pluto, its array of moons, and its surroundings. In doing so it will also perform the first exploration of the Kuiper Belt—the opening of an entirely new astronomical frontier.

Right now we know ridiculously little about Pluto. We know it has an atmosphere consisting largely of nitrogen, like our Earth's though drastically less dense. It has an ultra-cold crust covered with ices of nitrogen, carbon monoxide, and methane. It has at least five moons, polar caps, and an interior that is primarily composed—surprisingly—of rock. Most important, we know that Pluto is the archetype for an entire class of planets that have never been explored. Beyond that, it is a mystery, a virgin world. Who knows what discoveries await?

The great lesson of planetary exploration—from the 1960s flybys of Mars and Venus to the initial explorations of Mercury and Jupiter, Saturn, Uranus, and Neptune—is to expect the unexpected. No one expected dry riverbeds on Mars. No one expected Mercury to be an exposed planetary core with its mantle stripped away, or to find volcanoes and geysers on the moons of giant planets. No one expected oceans inside Jupiter's moon Europa, or ice in the clouds of Venus. All of these surprising truths emerged from the early reconnaissance missions.

As my team and I prepare for *New Horizons*'s encounter with Pluto, we are preparing to be surprised yet again by the richness of nature and the grandeur of seeing a new, faraway planet for the first time. *New Horizons* is a small spacecraft. It is dwarfed by *Voyager 1* and *2* that preceded it to open up the exploration of giant worlds, and it costs barely one fifth as much as the *Voyager* project. Nevertheless, it carries much more powerful scientific instruments. By analogy with the computing revolution we've witnessed since the 1970s when the *Voyagers* were built, *New Horizons* is like a tablet computer compared to Voyager's mainframe, packing much greater capability into a much smaller volume, and at a much lower price.

Beginning in May, *New Horizons* will deliver higher-resolution images of Pluto and its satellites than are possible from any telescope on Earth—even the Hubble Space Telescope. For 10 weeks before and after the day of encounter, it will "own" the system. At closest approach *New Horizons* will sample Pluto's atmosphere, search for new moons, look for possible rings, map the composition and temperature distribution across all the bodies in the Pluto system, and take images so good that if it were making an equivalent pass over New York City it could spot wharfs on the Hudson River.

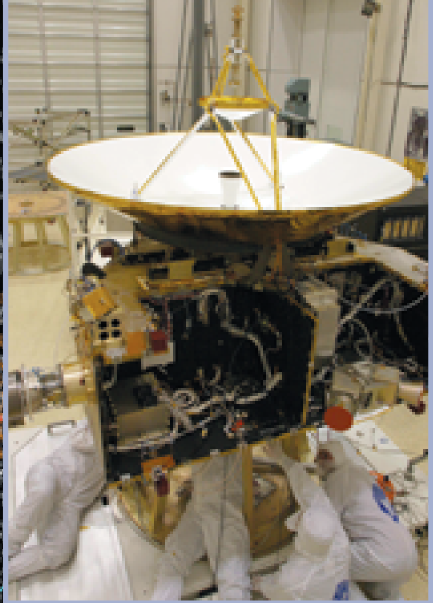
I have worked for 25 years to make the *New Horizons* mission happen, because the scientific promise is so great. The exploration of Pluto will mark both the opening of the exploration of the Solar System's third zone and the historic closing of the initial reconnaissance of our planetary system as a whole.

Where will you be when humankind makes its farthest-ever landfall? What will you tell your children and grandchildren you learned about space because you were there with *New Horizons*, riding along virtually on television or the Internet? And what will you tell them you learned about ourselves, this wonderful species that seeks to know the universe from which it was born?

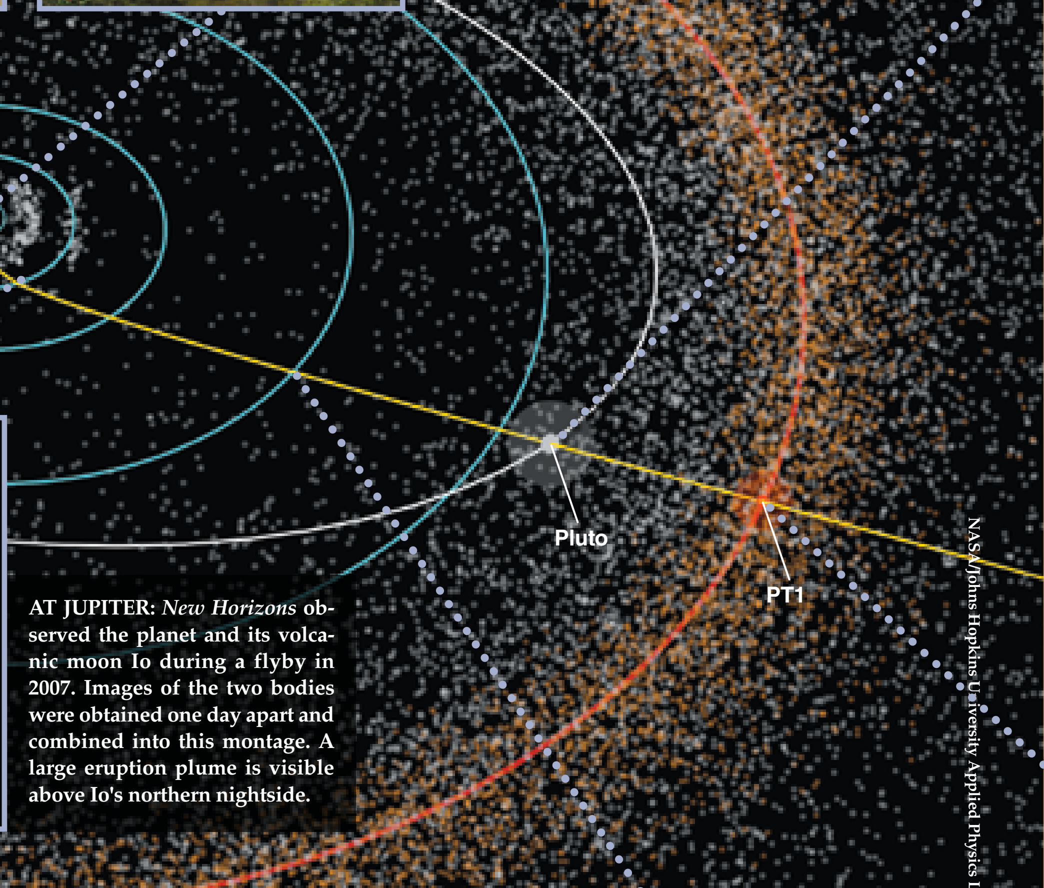
PREVIEW: Illustration of *New Horizons*'s flight past Pluto and its largest moon, Charon, is guided by paltry Earth-based observations. Pluto has strong markings and a thin nitrogen atmosphere. Almost every other detail will be a surprise.

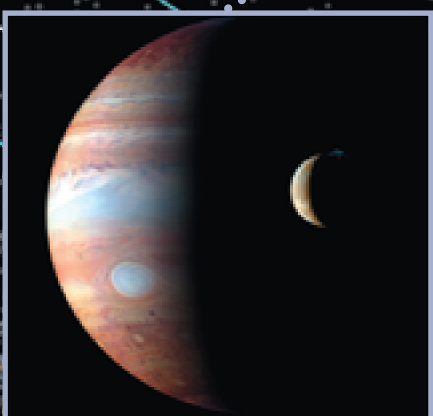
NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute/Steve Gribben

S. Alan Stern is a planetary scientist and the principal investigator of NASA's New Horizons *mission. He is former head of NASA's space and Earth science program and is slated to fly to space in 2016 as a researcher on both Virgin Galactic and XCOR suborbital spacecraft.*



BEGINNING: *New Horizons* team members performed a systems check on the 2.1-meter main antenna (*far left*) in February 2005, while the probe was under construction at Johns Hopkins University's Applied Physics Lab. Liftoff took place on January 19, 2006, from Cape Canaveral, Florida (*left*). Riding atop an Atlas V rocket, *New Horizons* became the fastest spacecraft ever launched; it passed the distance of the Moon in nine hours.





AT JUPITER: *New Horizons* observed the planet and its volcanic moon Io during a flyby in 2007. Images of the two bodies were obtained one day apart and combined into this montage. A large eruption plume is visible above Io's northern nightside.

NEW HORIZONS INSTRUMENTS

Ralph is the visible and infrared imager/spectrometer on *New Horizons*. It will provide color, composition, and thermal maps.

Alice is an ultraviolet imaging spectrometer. It will analyze the composition and structure of Pluto's atmosphere and look for atmospheres around Charon and any Kuiper Belt objects visited after the Pluto encounter.

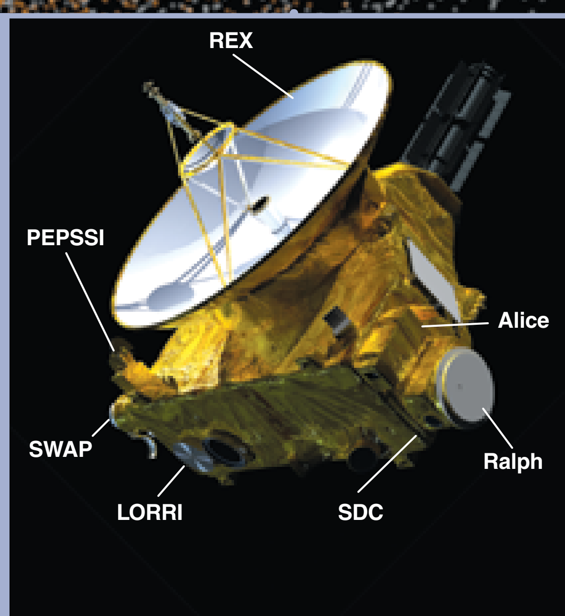
REX (Radio Science Experiment) will measure atmospheric composition and temperature by detecting distortions to radio signals from Earth.

LORRI (Long Range Reconnaissance Imager) is a telescopic camera that will map Pluto's far side and provide high resolution remote geologic data.

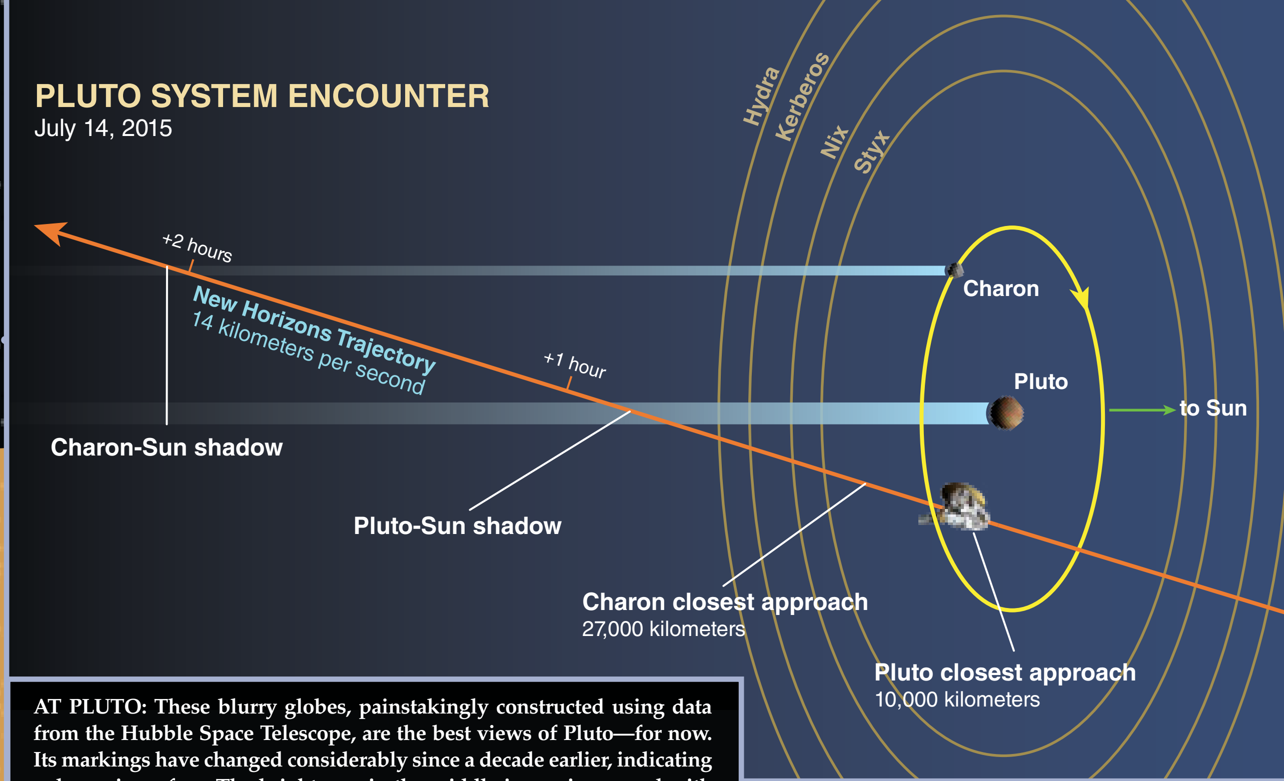
SWAP (Solar Wind Around Pluto) will measure the escape rate of Pluto's atmosphere and observe Pluto's interaction with the solar wind.

PEPSSI (Pluto Energetic Particle Spectrometer Science Investigation) will measure composition and density of ions escaping from Pluto's atmosphere.

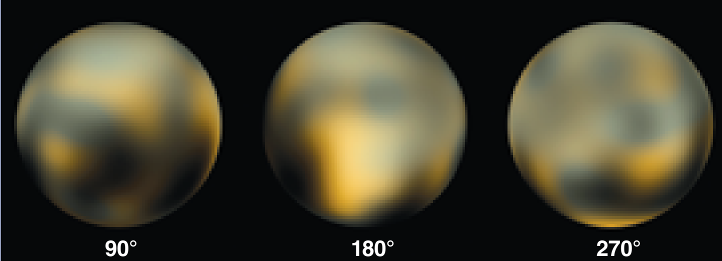
SDC (Student Dust Counter), built and operated by students, is measuring the space dust peppering *New Horizons* as it travels across the Solar System.

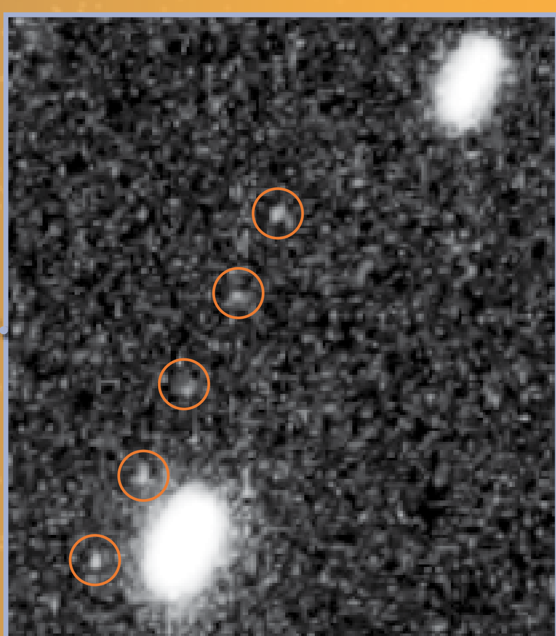


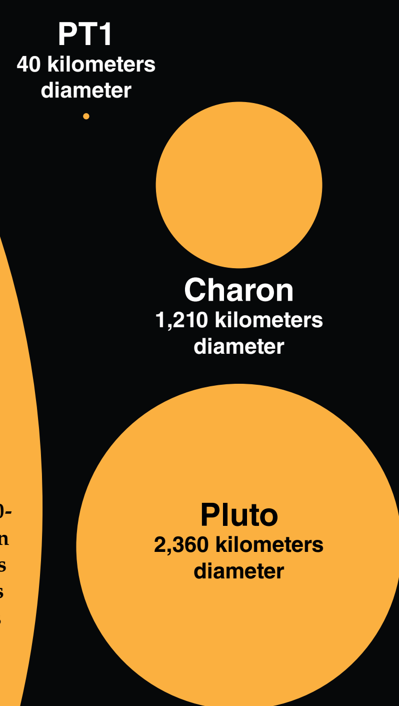
NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute/Alex Parker, Steve Gribben



AT PLUTO: These blurry globes, painstakingly constructed using data from the Hubble Space Telescope, are the best views of Pluto—for now. Its markings have changed considerably since a decade earlier, indicating a dynamic surface. The bright area in the middle image is covered with carbon monoxide frost; it will be a high-priority target for *New Horizons*.







Earth
12,742 kilometers
diameter

EXTENDED MISSION: A 40-kilometer-wide object known as PT1 could be *New Horizons*'s next stop after Pluto (PT1 stands for "potential target 1"). It was discovered by the Hubble Space Telescope during a dedicated search for a follow-on destination; it was identified by its motion (*red circles*) relative to the stars. Pluto is the largest member of the Kuiper Belt, the zone of icy bodies beyond Neptune. Little PT1 is more typical of the myriad objects out there. It is probably unchanged since the birth of the Solar System.

RELATIVE SIZE: Pluto is small compared to Earth. Its size and location within the Kuiper Belt are why the International Astronomical Union reclassified it as a "dwarf planet," but Pluto is enormous compared to most other Kuiper Belt Objects, such as PT1. Pluto also has a unique relationship with Charon, which is by far the largest moon relative to its parent planet. Pluto is the key to understanding the outer Solar System and its connection to the evolution of Earth and the other planets.

Discovering the Edge of the Solar System

Recent discoveries suggest that planets larger than Pluto may exist in the outer reaches of our solar system

Chadwick A. Trujillo

The last planet to be discovered in our solar system was Pluto, found by Clyde Tombaugh nearly 75 years ago. Pluto lay just beyond the most distant gas giants, Uranus and Neptune, and its discovery suggested to astronomers that the edge of the solar system had yet to be found. What's more, Pluto has some strange qualities compared with its giant neighbors: It's about a thousand times less massive, it's made mostly of rock and ice (as opposed to gas), and it has an elongated orbit, which is inclined to the orbital plane of the solar system. This oddity was an early clue that the outer limits of the solar system might be worlds apart in character from the giant outer planets.

Although various scientists had suggested that the region beyond Neptune consists of more than just Pluto, it wasn't until 1992 that another trans-Neptunian object was discovered by David Jewitt and Jane Luu, both then at the University of Hawaii. The object they found, with the unassuming name of (15760) 1992 QB_1, appears to be about 130 kilometers in diameter, giving it the same land area as West Virginia. This is tiny compared with Uranus and Neptune, which are each about 50,000 kilometers across, and still much smaller than Pluto, which has a diameter of about 2,320 kilometers.

Chadwick A. Trujillo is a postdoctoral fellow at the California Institute of Technology. He has been involved in the discovery of more than 100 Kuiper belt objects, including the largest and brightest object so far, Quaoar. His research interests include the Kuiper belt, solar system sciences, extrasolar planets and the applications of astronomical adaptive optics. In his spare time, he and his wife Tara hike, bike, swim and practice yoga. Address: California Institute of Technology, Division of Geological and Planetary Sciences, MS 150–21, Pasadena, CA 91125. Internet: chad@gps.caltech.edu

Before (15760) 1992 QB_1 and similar trans-Neptunian objects were found, scientists generally believed that these bodies would be between 1 and 10 kilometers across. But since 1992 more than 800 objects have been discovered, and most of them are about 100 kilometers in diameter. Together these trans-Neptunian bodies form a relatively flat, ring-shaped structure around the Sun that has been dubbed the "Kuiper belt," in recognition of Dutch astronomer Gerard Kuiper's 1951 proposal that undiscovered bodies lay outside the orbits of Neptune and Pluto. It turns out that an Irish amateur astronomer, Kenneth Edgeworth, was the first to mention the existence of such a region in 1943, so it's also sometimes called the "Edgeworth-Kuiper belt." (In 1930, just six months after Pluto's discovery, University of Chicago astronomer Frederick Leonard also mentioned the possibility that more Pluto-sized bodies might exist, but the name "Leonard-Edgeworth-Kuiper" has not received much attention.)

Edgeworth and Kuiper originally proposed the idea of a trans-Neptunian belt to help explain the origin of short-period comets, those with orbital periods of less than 200 years. In this view the belt served as a reservoir for the icy objects that would occasionally fall into the inner solar system. It is now generally recognized that this is indeed the case. However, the Kuiper belt is even more interesting because the bodies are considered to be the primordial remnants of the nascent solar system. The Kuiper belt objects, or KBOs, are some of the oldest unaltered bodies in the solar system and so represent an unexplored source of knowledge about its early history.

Interestingly, the KBOs seem to be even larger than many scientists suspected before the first KBOs were found. In June 2002, my colleague Michael Brown at the California Institute of Technology and I discovered the largest known KBO, which we named Quaoar (pronounced "KWAH-o-wahr"), after a deity of the native American Tongva people, many of whom still live in the Los Angeles area. Quaoar is about 1,300 kilometers in diameter, roughly half the size of Pluto. The existence of such a large KBO, along with a few others in the 1,000-kilometer range, suggests that there may be even larger objects out there. In this article I briefly describe what scientists have learned about the KBOs, especially the very large ones that were discovered in the past few years.

The Kuiper Belt Today

The Kuiper belt has only been known for 11 years, but its overall structure is reasonably well understood, even if a number of mysteries remain. The 800 known KBOs constitute only about three percent of the expected total population, which can be estimated by extrapolating from the very small portion of the sky that has been searched. Using such methods astronomers have surmised that there are about 30,000 KBOs larger than 100 kilometers in diameter. This makes the Kuiper belt population about 100 times greater in number and mass than the asteroid belt between Mars and Jupiter.

There appear to be three types of KBOs, which are distinguished primarily by their dynamic properties. About 50 percent of the KBOs have nearly circular orbits (eccentricities less than 0.2) and inhabit a ring-shaped structure between 42 and 48 astronomical units (AU) from the Sun (one AU is the distance between the Sun and the Earth). These are called the classical

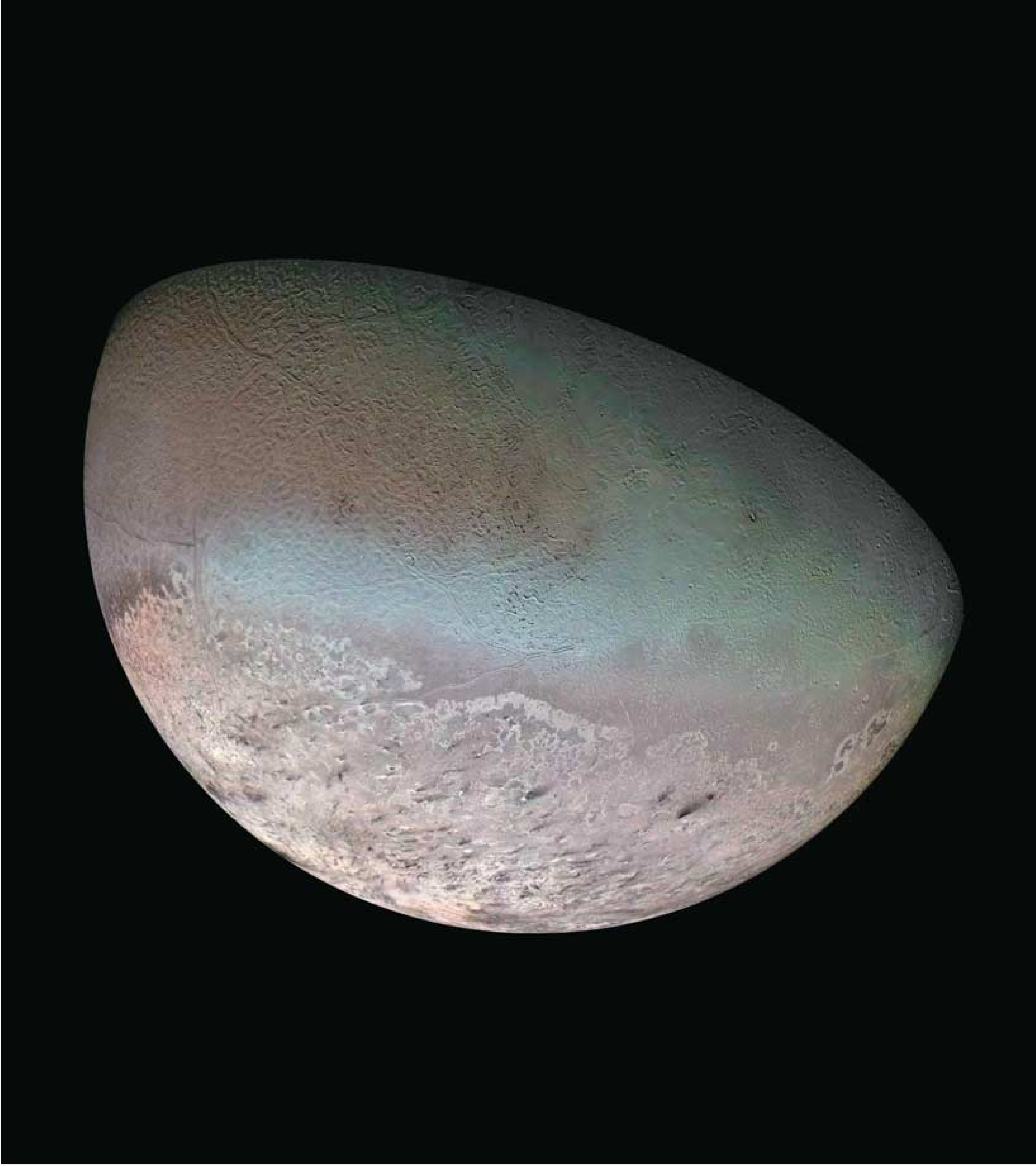


Figure 1. Surface of Neptune's largest moon, Triton, offers a clue to the appearance of the large objects of the Kuiper belt, a ring of minor planets marking the edge of our solar system. The cantaloupe-like texture may indicate large-scale geologic processes, whereas the small dark spots may be "cryovolcanoes" with liquid nitrogen bubbling through the solid nitrogen on the surface. The odd coloring is probably caused by simple organic compounds. To date more than 800 objects have been found in the Kuiper belt. It's believed that there are more than 30,000 Kuiper belt objects (KBOs) with diameters of about 100 kilometers across, a few that are about 1,000 kilometers across and one or more that may be larger than Triton (2,700 kilometers across). The image was made by the Voyager II spacecraft in 1989. (Image courtesy of NASA.)

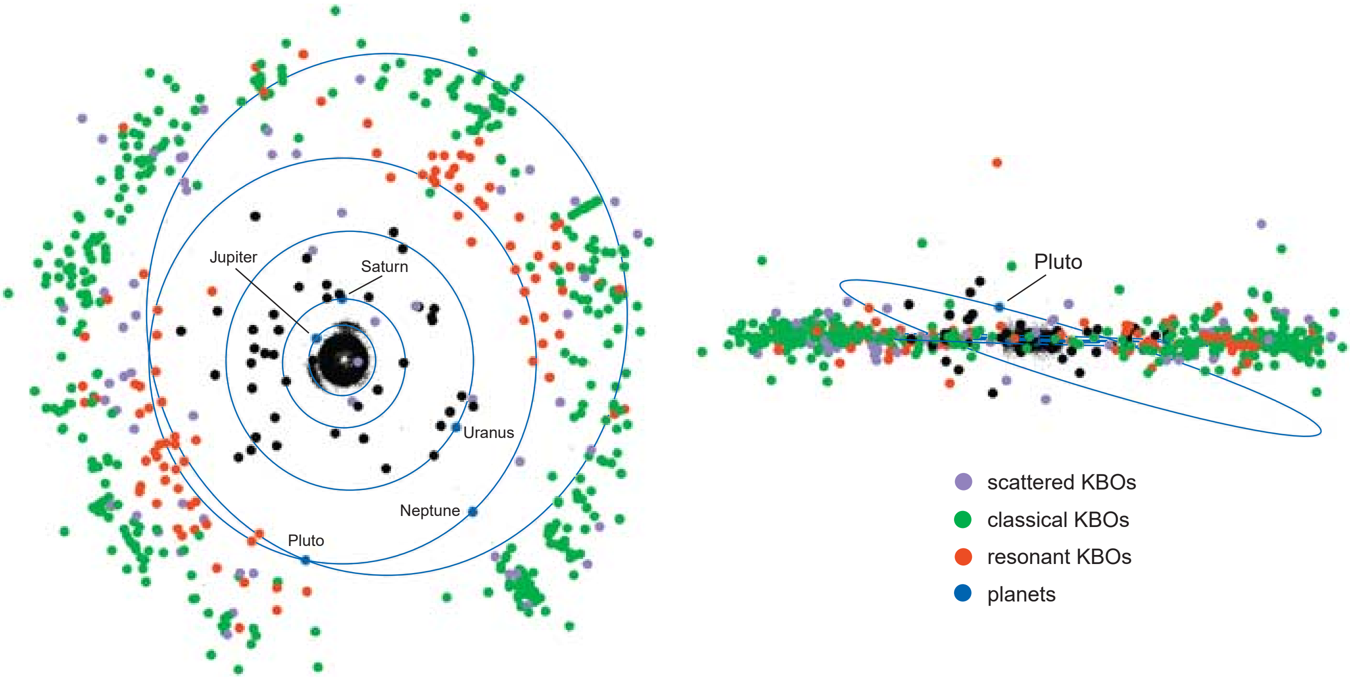


Figure 2. Kuiper belt objects generally lie outside the orbits of Neptune and Pluto, the outermost planets in our solar system *(left).* There are three classes of KBOs: "classical" *(green dots),* "resonant" *(red dots)* and "scattered" *(purple dots),* which are characterized by their dynamic properties *(see text).* The other minor planets in our solar system reside in regions closer to the Sun: The asteroid belt *(dense black stippling)* inhabits the region between Mars and Jupiter, Trojan asteroids flank Jupiter along its orbital path, and the Centaurs *(black disks)* have unstable orbits in the vicinity of the gas-giant planets. An edge-on view of the Kuiper belt *(right)* reveals that it is much thicker than predicted by planet-formation models, which propose a paper-thin disk much like Saturn's rings. This suggests that the Kuiper belt may have been gravitationally perturbed since its formation, increasing the speed of the objects and "puffing up" the disk. Note also that Pluto's peculiar orbit *(right, blue line)* is tilted, which is similar to the orbits of many resonant KBOs.

Ronald Royer/Science Photo Library

Figure 3. Short-period comets—those with orbital periods of less than 200 years—are believed to originate in the Kuiper belt. KBOs are probably thrown inward by chaotic interactions with Neptune; they often break into smaller pieces and begin outgassing as they approach the Sun. Photographed on September 4, 1989, comet Brorsen-Metcalf was first observed in July 1847 and has a period of about 71 years.

KBOs because they more or less conform to the original idea that the Kuiper belt consisted of a relatively thin disk at the edge of the solar system. Another 10 percent of the KBOs are in mean-motion resonance with Neptune (as is Pluto). Most of the resonant KBOs orbit the Sun only once or twice for every three times Neptune orbits the Sun. Resonance protects these small, icy KBOs from close passage to the giant planet, which would otherwise pull them towards the inner solar system (where they might become comets), or fling them even farther from the Sun. The remaining 40 percent appear to have undergone a weak gravitational interaction with Neptune, which has scattered them into highly eccentric orbits. The orbits of the scattered KBOs can carry them anywhere from the inner part of the classical belt outward to the most distant parts of the known solar system, up to 30 times the distance between the Sun and Neptune, roughly 900 AU!

Although these three populations have been identified, there are some fundamental observations about the belt that currently defy explanation. Among these is the presence of the out-

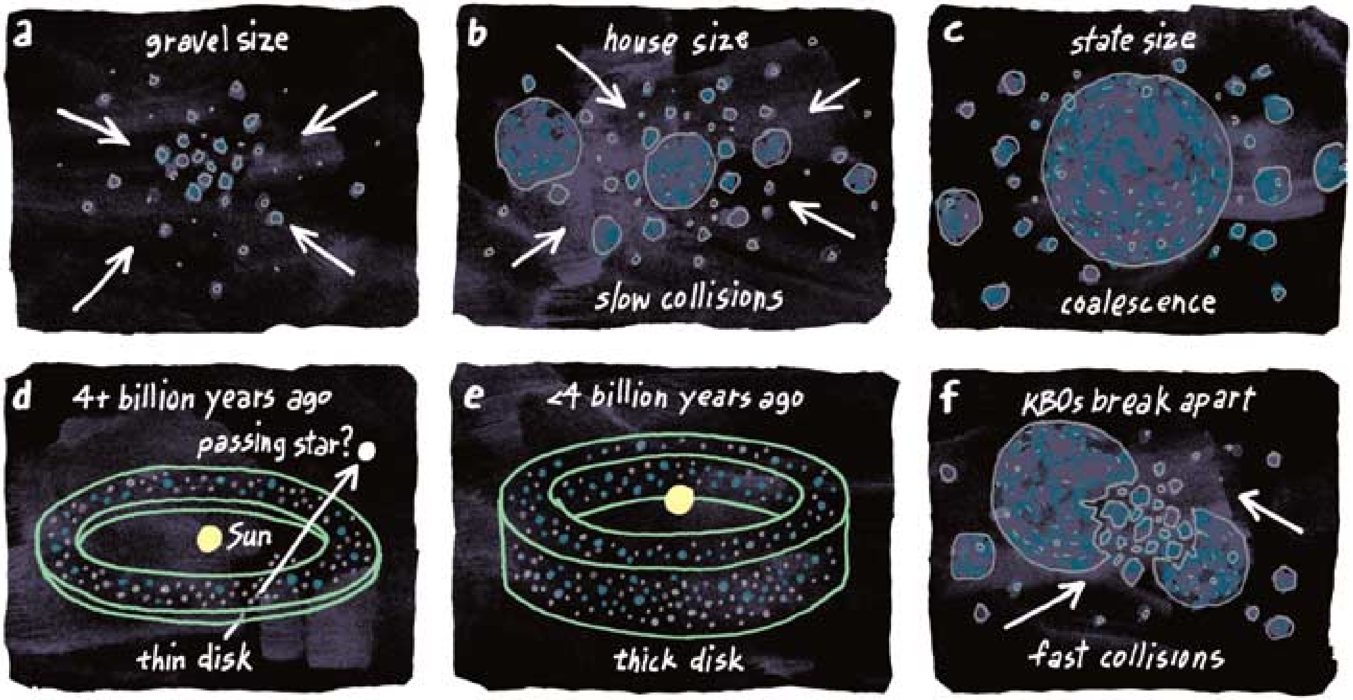


Figure 4. The Kuiper belt formed more than four billion years ago when the solar system was just taking shape. Slow-moving gravel-sized debris at the solar system's edge gradually coalesced through gravitational attraction, eventually forming objects with a land area equal to large American states *(a, b and c)*. Early in its history the Kuiper belt was a flattened disk *(d)*, but later something—perhaps a passing star—disrupted the disk, accelerated the KBOs and sent them into more highly inclined orbits. Today the Kuiper belt is relatively thick *(e)*, and when the fast-moving KBOs collide they break into smaller pieces *(f)*.

er edge displayed by the classical Kuiper belt: Beyond 48 AU none of the KBOs have circular orbits. There are several possible explanations for this outer edge, all involving catastrophic events during the early stages of our solar system's formation. In one scenario, material in the outer solar system was stripped away by a close-passing star. In another view, Earth-sized planets may have passed through the primordial Kuiper belt, disrupting its disk structure. A third possibility is that some additional resonances with Neptune may have played a role in creating the outer edge. However, no theory has been widely accepted.

The Kuiper belt's thickness also came as a surprise to observers. The basic theory was that KBOs inhabited a very thin disk, less than one degree above or below the plane of the solar system—something that was required for the KBOs to form by gravitational accretion. In fact, the Kuiper belt is actually about 10-degrees thick, plus a "halo" of objects with much higher inclinations. This extra thickness suggests that the belt was somehow stirred up after the solar system formed, possibly by some of the same processes suggested to explain the outer edge.

The KBOs vary in color from nearly the reddest objects in the solar system to a bland gray. Although much work has been done to measure the colors of the KBOs, there are few definite conclusions. It does appear that KBOs that travel in very circular orbits in the plane of the solar system tend to be red. Some scientists have suggested that collisions could be responsible for the coloring. Left undisturbed, KBOs will slowly redden over time as they are "baked" by galactic cosmic rays and high-energy ultraviolet radiation from the Sun. Collisions could excavate pristine gray material from the interior. So objects that have suffered recent collisions might appear gray, whereas objects with older surfaces would appear darker and redder. Others have proposed that the observed color distribution is a leftover from the formation of the solar system—the different colors corresponding to regional differences in the chemistry and composition of the primordial solar nebula. It's known that the chemicals that dominate the planetary bodies vary greatly: The inner planets are mostly rock, whereas the outer "gas giants" are made of gas and ice. For example, if some of the KBOs formed near Jupiter and later moved beyond Neptune, while others formed beyond Neptune and stayed there, one might expect to see differences in color and composition between these two populations.

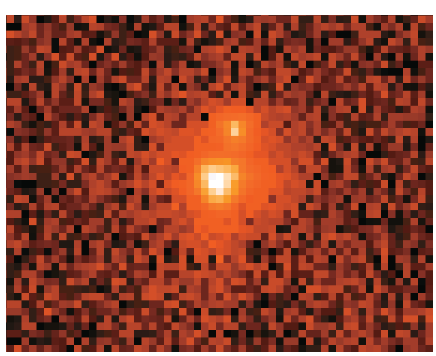


Figure 5. Some Kuiper belt objects have moons—a surprising discovery. Traditional mechanisms to explain the formation of moons, such as collisions or capture, do not work well for the binary KBOs because of their small masses and large separations. Solving the mystery of how these binary systems formed should provide clues to the Kuiper belt's early environment. This pair—KBO 1999 TC_{36} and its moon—are separated by about 8,300 kilometers, about one-fiftieth the distance between the Earth and its moon. (Unless noted, all photographs courtesy of the author.)

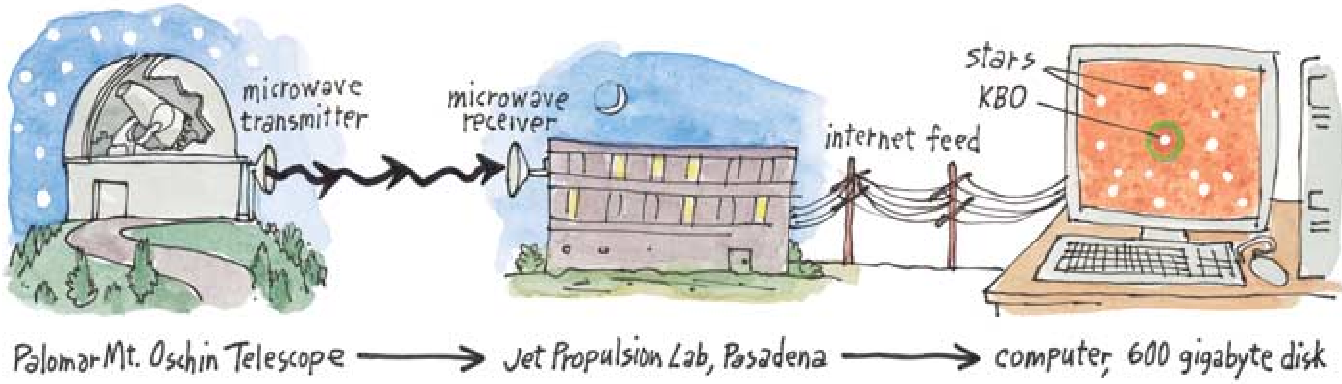


Figure 6. Semi-automated search procedures and robotic telescopes have allowed planetary scientists to search for KBOs from their office desktop computers. Robotic telescopes, such as the Palomar Oschin Telescope, can systematically scan the sky with a relatively small mirror (1.2 meters) and record images with a charge-coupled device, or CCD camera *(see Figure 7)*. This method has rapidly increased the pace of discovery in the past few years.

We know very little about the surfaces of the large KBOs, except that their colors vary widely and that some appear to have water ice (as revealed in spectroscopic studies). This isn't much of a surprise because the KBOs are thought to be the parent bodies of the comets, which are rich in water ice. Unfortunately, the vast majority of telescopic studies of KBOs have revealed very little information about their surface composition, mostly because even the brightest of the objects are difficult to study from Earth. One of the best clues to the appearance of a KBO's surface was provided in 1989 when Voyager II cruised by Neptune's large moon Triton—several years before the first KBO was discovered.

In fact, Pluto, its moon Charon, Triton and the Kuiper belt may be more closely related than one might guess at first. Many scientists have considered the possibility that Pluto may be accurately characterized as the largest KBO rather than a planet. (The Rose Center for Earth and Space in New York city, for example, removed Pluto from its list of planets in 2000.) This is because there are about 10 or 20 KBOs that have orbits very similar to Pluto's. Pluto's moon, Charon, was probably formed during a collision between Pluto and a KBO in the early solar system—making Charon a product of the KBOs. Triton, because it circles Neptune "backward" (compared with the orbits of Neptune's other moons and with the orbits of the planets around the Sun, is believed to be a captured satellite—a KBO-turned-Neptunian moon. Indeed Triton is often suggested as a possible analogue to Pluto's surface, which will be explored when NASA's *New Horizons* Pluto-Kuiper belt mission visits the small planet in 2015.

Several KBOs also appear to have their own little moons. These are often very far from the parent body—between 10,000 and 100,000 kilometers—and quite large, nearly one-third the diameter of the parent body. Their separations relative to their sizes are somewhat similar to the Earth-Moon system, which is quite surprising because the KBOs are typically about one-millionth the mass of the Earth. Their existence is quite inexplicable by the standard collisional mechanism used to explain our Moon's formation—the impact of a Mars-sized body on the Earth.

This assortment of mysteries is particularly vexing because it's not clear whether an experiment could be devised that would lead to a definitive explanation of the observations. Most of the belt's features are probably relics of the solar system's formation, and so far we know little about what happened in the outer solar system. Theoreticians attempt to model this process with dynamic simulations of growing bodies in the Kuiper belt. The standard paradigm is that the KBOs formed from the coalescence of smaller bodies,

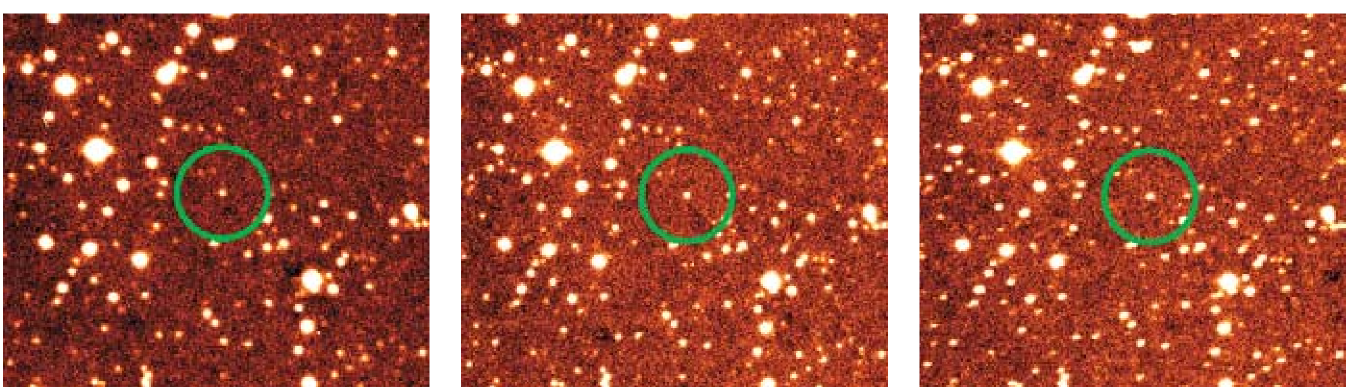


Figure 7. Discovery of the largest known KBO—called Quaoar—involved a comparison of three CCD images taken just a few hours apart. Computers scan images of the sky, searching for moving objects against the fixed background of stars, but a human being is necessary to confirm the discoveries. Each image represents only a small fraction of the sky (perhaps 1/3,500 or less) that is imaged every night. Here Quaoar *(circled)* is moving left to right relative to the stars. (Hint: Note the subtle "movement" of the star at four o'clock on the green circle.)

perhaps one kilometer across, early in the solar system's history. At some point, however, there was a drastic change in the properties of the Kuiper belt, which arrested the gravitational growth of the individual KBOs. This may have been the event or series of events that stirred up the belt to make it thicker and produced the outer edge.

We can, in fact, make an observation that can tell us how far the growth process proceeded—that is, we can look for the largest KBO. This would tell us how large a body could grow before the aggregation stopped. But this is not a small task; it means we must look through the entire sky and find the largest KBOs.

Searching for KBOs

The biggest KBOs are also the brightest, so very large telescopes are not needed. However, the search is still difficult for the simple reason that the sky is immense. To get some sense of the task, it helps to know that a typical "wide-field" professional telescope capable of detecting the brightest KBOs has a field of view that is about one-third the diameter of the full Moon, or about 0.2 degree across. It takes about 150,000 full Moons to fill the sky, so if one took a wide-field exposure every minute, it would require nearly six years of clear nights to search the whole sky. My colleagues and I are currently engaged in the task of searching for large KBOs by using a telescope with an extremely wide field of view and a camera that captures nearly three square degrees, about 10 times the area of the full Moon.

The camera is a charge-coupled device, or CCD, a silicon-based electronic instrument that can capture faint astronomical images in a fraction of the time needed for film or other emulsion-based systems. These cameras have three main advantages over conventional photography: They are about 100 times more sensitive to photons, they are very sensitive to red light (where KBOs are brightest), and the images collected during a CCD exposure are downloaded directly to a computer, immediately facilitating processing.

We use a semi-automated procedure (some human intervention is involved), and this has had a profound impact on the search for KBOs simply because a machine is so much faster (and less prone to error) than a human being. Many large KBOs, including Quaoar,

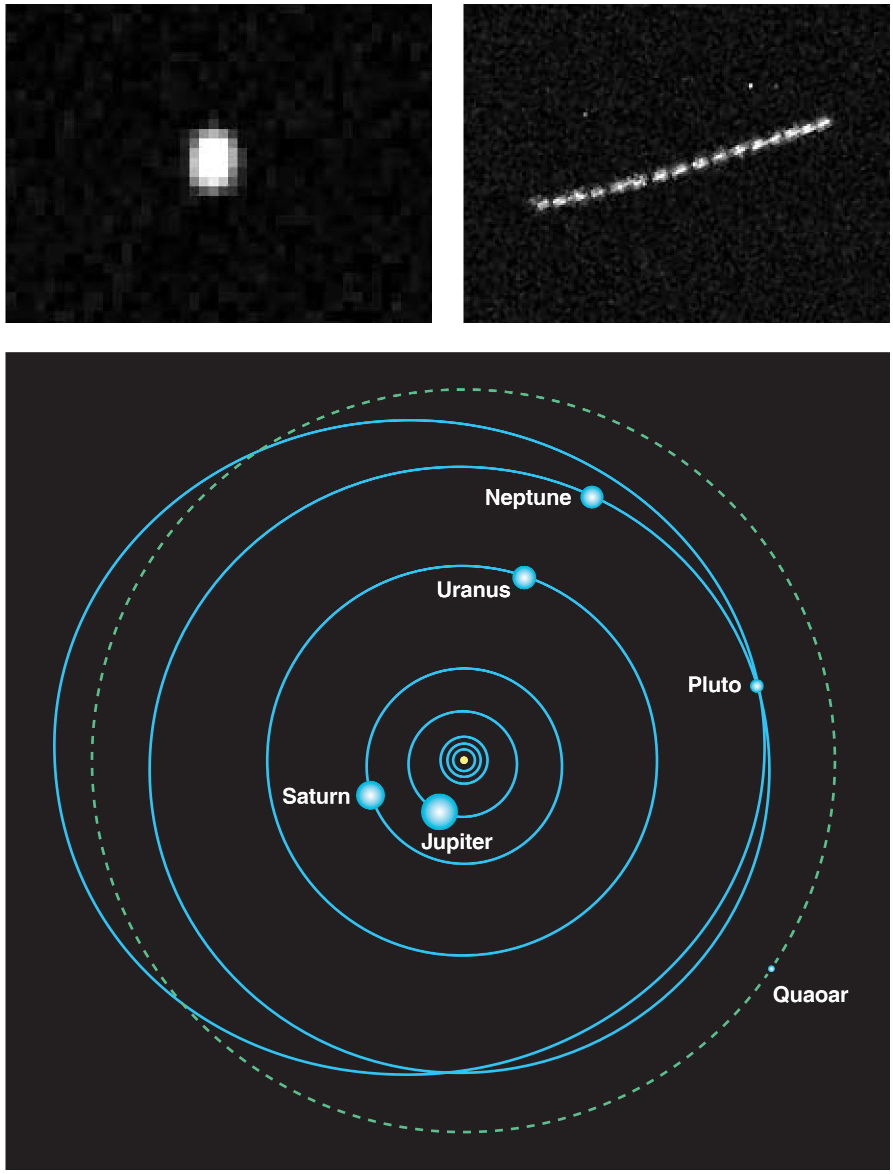


Figure 8. Quaoar ***(top, left)*** **is the most distant object in the solar system ever to be resolved by a telescope—about 6.5 billion kilometers from Earth. The image, made by the Hubble Space Telescope, led to a direct measurement of the object's diameter: 1,250 kilometers. Sixteen snapshots** ***(top, right)*** **of Quaoar were taken on July 5, 2002, as it moved across the sky over a period of 29 minutes. Quaoar has one of the most circular orbits known in the solar system** ***(bottom).*** **(Top two images courtesy of NASA and Michael Brown, Caltech.)**

have been discovered this way. Three images of the same patch of sky are taken in sequence, about 90 minutes apart. Each exposure is only about 2.5 minutes long, which is much shorter than the many nights of exposure required to reveal the most distant galaxies. Each image is analyzed by an automated algorithm, which identifies objects that appear to move during the time sequence. Many false positives are identified along with the real moving objects, and so an astronomer must look at the images to eliminate the "false discoveries." This semi-automated approach is also used by many astronomers searching for potentially hazardous near-Earth asteroids.

In the past few years, the search process has also been aided by the development of robotic telescopes. These telescopes have the ability to operate autonomously, examining a small part of the sky and then systematically moving on to the next target. Most are multi-million-dollar opto-mechanical systems, coupled to millions of dollars of detector instrumentation, all housed in a protective enclosure. Even though some telescopes have the ability to

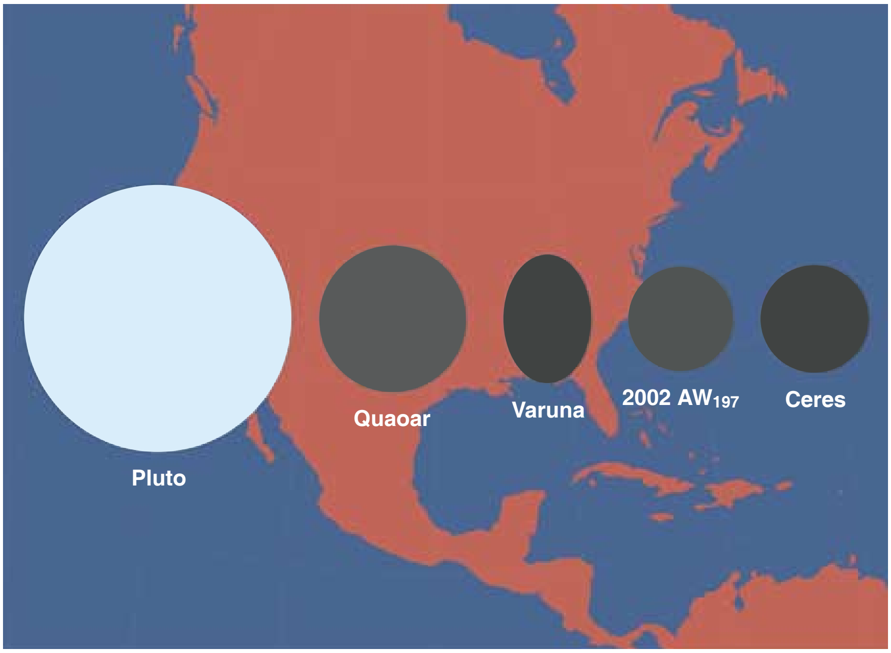


Figure 9. Pluto and the four largest minor planets could sit atop the North American continent. Pluto is the smallest planet and is often considered to be the largest Kuiper belt object. Quaoar and 2002 AW_{197} were discovered by the author and his colleagues during their survey for bright KBOs. Varuna is a large KBO that spins so quickly its shape has been distorted. Ceres, discovered in 1801, is the largest object in the asteroid belt between Mars and Jupiter. The shading represents the relative surface brightness of the objects; Pluto appears relatively bright because of nitrogen frost on its surface. There may be as many as 10 similarly large undiscovered KBOs in the solar system.

close automatically during hazardous weather, they are almost always monitored by a human operator nearby.

Our ongoing survey of the sky for KBOs uses the Palomar Oschin Telescope, which has a mirror diameter of 1.2 meters—relatively tiny compared with the 8 to 10-meter giants that have been built in the past decade. However, the Oschin telescope is a reliable workhorse that has been in operation since 1948. Fortunately for us it was recently upgraded by the Near Earth Asteroid Tracking program (NEAT), which searches for asteroids that might collide with the Earth. This telescope executes its exposures robotically, seamlessly trading time between our survey and the NEAT program.

The combination of a semi-automated telescope and analysis software allows large-scale surveys to be done with a minimum of manpower. Our survey, which in a year and a half has covered about 5,000 square degrees (roughly one-eighth of the sky), can be operated by a single astronomer working during the day and a small crew to ensure proper maintenance and operation of the telescope. Clyde Tombaugh spent 14 years examining 30,000 square degrees of photographic plates by eye to find Pluto. At our current pace we'll be able to cover the entire sky in about 12 years, with instruments that are 50 times more sensitive than he used—a striking testimony to the advantages of silicon-based technologies.

Quaoar and the Big KBOs

Quaoar is about 100 times brighter (at red wavelengths) than the typical KBO and about twice as bright as the next brightest object. So why wasn't Quaoar discovered before June 2002? Quaoar is just a little too dim for Tombaugh to have discovered it, but one would have expected that other surveys in the 20th century might have seen it. And, indeed, a look at previous sky surveys reveals that others had recorded its presence, but did not recognize it because they weren't looking for moving objects. Those who were looking for moving objects weren't looking for ones that were moving so slowly (and, hence, were so far away). In some instances Quaoar was near the survey's faintness limit. Although Quaoar is considered to be bright, it is still about 100 times fainter than Pluto, which is considered to be a very difficult target for eyeball-based observing even with large amateur telescopes.

Determining the size of Quaoar is no easy task. The telescope records an object's position and brightness, but one must fit an orbit to at least three observed positions to produce a distance estimate. We see Quaoar by light that has traveled 6.5 billion kilometers from the Sun to Quaoar's surface and then reflected back to the Earth another 6.3 billion kilometers. By knowing Quaoar's distance and the amount of light reflected back to a telescope, we can get a general idea of its size. But

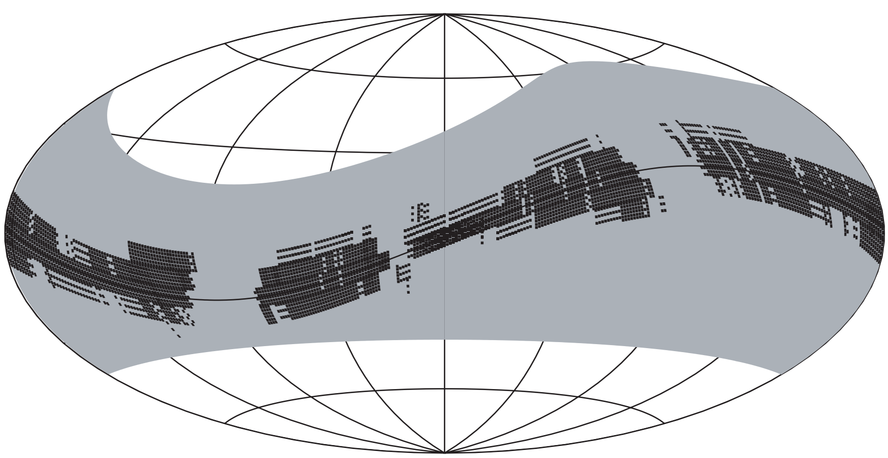


Figure 10. Clyde Tombaugh's photographic plate survey between 1929 and 1945 covered about 70 percent of the sky *(gray)* and resulted in the discovery of Pluto. It's been estimated that Tombaugh spent 7,000 hours scanning the plates for moving objects. Pluto was his only discovery at the edge of the solar system. The bright KBO survey undertaken by the author and his colleagues since 2001 has covered much less sky *(black)*, but can search for objects 100 times fainter than Pluto, resulting in the discovery of 30 KBOs to date. The ecliptic *(black line)* marks the plane of the solar system, where the concentration of KBOs is greatest.

because we don't know Quaoar's albedo—a small, snow-covered surface could reflect the same amount of light as a large, charcoal-covered surface—this technique is relatively crude.

One way to resolve this ambiguity is to measure the heat emanating from Quaoar. An object's thermal emission tells us how much light it has absorbed and so how dark its surface must be. These measurements were performed by a 30-meter radio telescope in Spain, at the Institute for Radioastronomy in the Millimeter Range (IRAM). The measurement is difficult to make because KBOs are very cold (about 42 kelvins)—only the five or so largest KBOs can be measured with any precision this way. By combining them with our own optical measurements of its position and brightness, we estimated Quaoar to be about 1,300 kilometers in diameter.

Although there are many systematic biases to this technique, we had an independent way to verify this measurement. It turns out that Quaoar is sufficiently large for the Hubble Space Telescope (HST) to resolve it into a disk. By comparing its apparent size (about 1.5 pixels) to that of a star (which is essentially a pinpoint, zero pixels across) we were able to determine that Quaoar is about 1,250 kilometers across. Such a precise measurement is difficult to perform because the optical system of HST must be very well understood. The comparison of the star is essential to demonstrate that we understand the distortions created in the HST optics, even though, after the corrective COSTAR installation in 1993, it is one of the most optically perfect telescopes in the world.

Where are the Super-Plutos?

The presence of the largest and brightest KBOs has only been confirmed in the past few years. Of the 10 intrinsically brightest KBOs, all have been found since 2000. Because only a small fraction of the sky has been examined for these bright KBOs, it is likely that more exist and remain to be discovered. Although our survey has found what is currently the largest KBO since Pluto's discovery, we have only covered a fraction of Tombaugh's survey, and an even smaller fraction of the whole sky. We hope to continue our survey, eventually covering about the same amount of sky as Tombaugh, but we should be able to detect objects 50 times fainter. We have only imaged about one-eighth of the whole sky, and about one-fifth of the sky where KBOs are most likely to be found (near the plane of the solar system). To date we have found about 30 bright KBOs, with Quaoar and 2002 AW_{197} having the largest diameters, about 1,250 and 900 kilometers, respectively.

As we search the remaining four-fifths of the sky, we expect to find a total of about 100 bright KBOs, with perhaps 10 KBOs in the 1,000-kilometer-diameter range. The size distribution of the Kuiper belt is such that for every 15 objects found of a given diameter, one will be found with twice that diameter. So of the 10 KBOs in the 1,000-kilometer range, it is possible that one will be about 2,000 kilometers across, approaching Pluto's size. Our survey should be completed in 2005, and with some luck one of the Pluto-sized bodies may appear in our data.

Another possibility is finding a very large object, perhaps as big as Mars. The sky has not been adequately searched for Mars-sized bodies at extreme distances from the Sun—twice as far as the main body of the Kuiper belt. If there are one or two such "super-Plutos" out there, they could easily have escaped Tombaugh's survey because of their faintness and all the other KBO-sensitive surveys because of the small amount of sky examined. There are no good theoretical or observational limits to the existence or nonexistence of such large bodies—except for the distant hope that we have yet to find the outer edge of our solar system.

Acknowledgments

The author wishes to thank Michael E. Brown, who provided valuable input to much of the research presented in this article.

Bibliography

Jewitt, D., and J. Luu. 1993. Discovery of the candidate Kuiper belt object 1992 QB_1. *Nature* 362:730–732.

Kowal, C. T. 1989. A solar system survey. *Icarus* 77:118–123.

Kuiper, G. P., and B. M. Middlehurst. 1961. "Planets and satellites." In *The Solar System*, ed. G. P. Kuiper and B. M. Middlehurst. Chicago: University of Chicago Press.

Tombaugh, C. W. 1946. The search for the ninth planet, Pluto. *Leaflet of the Astronomical Society of the Pacific* 5:73.

Trujillo, C. A., and M. E. Brown. 2001. The radial distribution of the Kuiper belt. *The Astrophysical Journal Letters* 554:L95–L98.

Trujillo, C. A., D. Jewitt and J. Luu. 2001. Properties of the trans-Neptunian belt: statistics from the Canada-France-Hawaii Telescope survey. *The Astronomical Journal* 112:457–473.

Trujillo, C. A., J. Luu, A. Bosh and J. Elliot. 2001. Large bodies in the Kuiper belt. *The Astronomical Journal* 122:2740–2748.

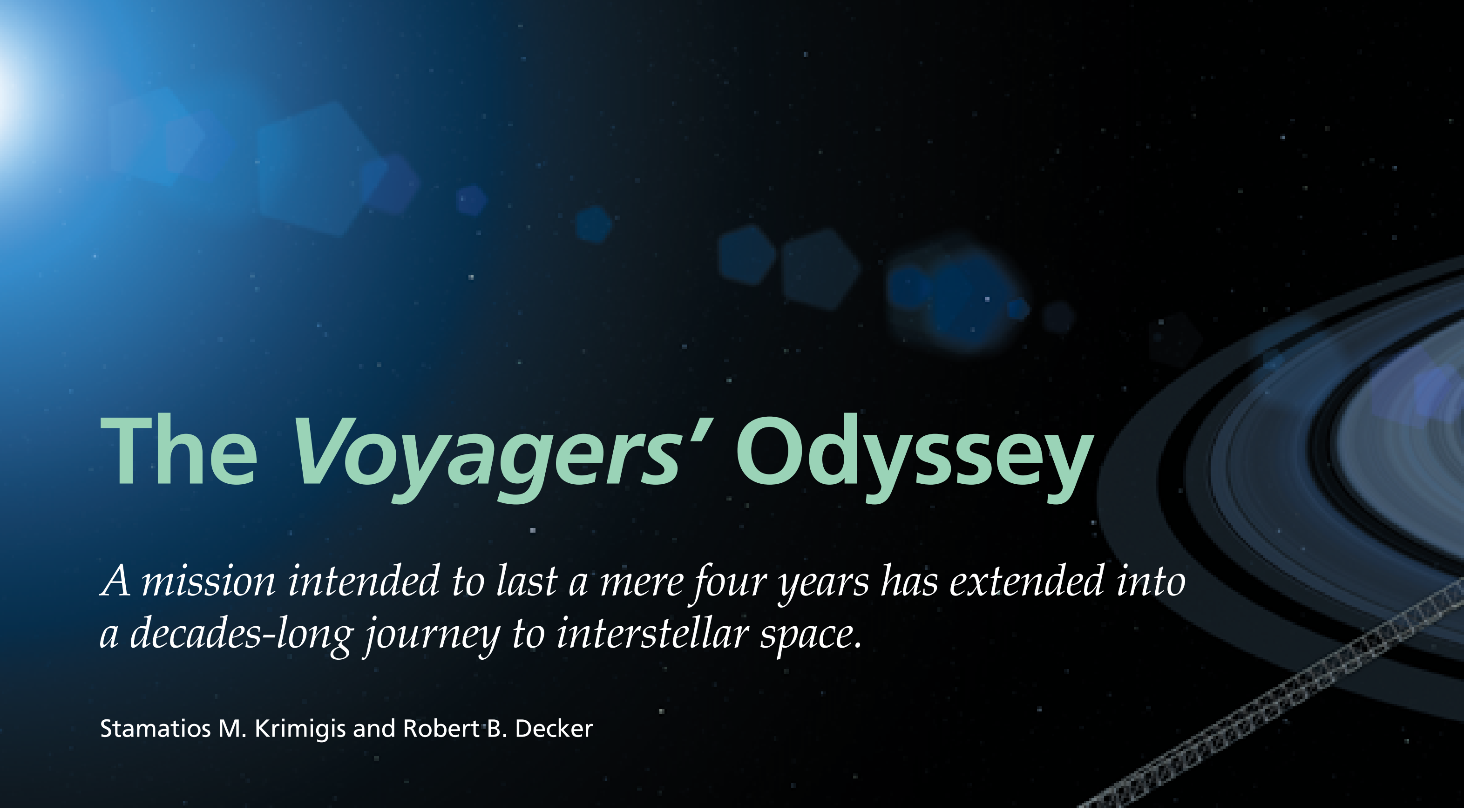
Williams, G. V., et al. 2002. 2002 LM60. *Minor Planets Circulars*, 34.

For relevant Web links, consult this issue of *American Scientist Online:*

http://www.americanscientist.org/template/IssueTOC/issue/400

I appreciate the diversity of articles and their accessibility to non-experts. I learn a lot about science going on outside my discipline (chemistry) and also about the history and social issues associated with doing science (recent article about women's contributions to Nobel Prize).

I simply enjoy the breadth of subjects covered by *American Scientist* and the thoroughness of treatment.



The *Voyagers'* Odyssey

A mission intended to last a mere four years has extended into a decades-long journey to interstellar space.

Stamatios M. Krimigis and Robert B. Decker

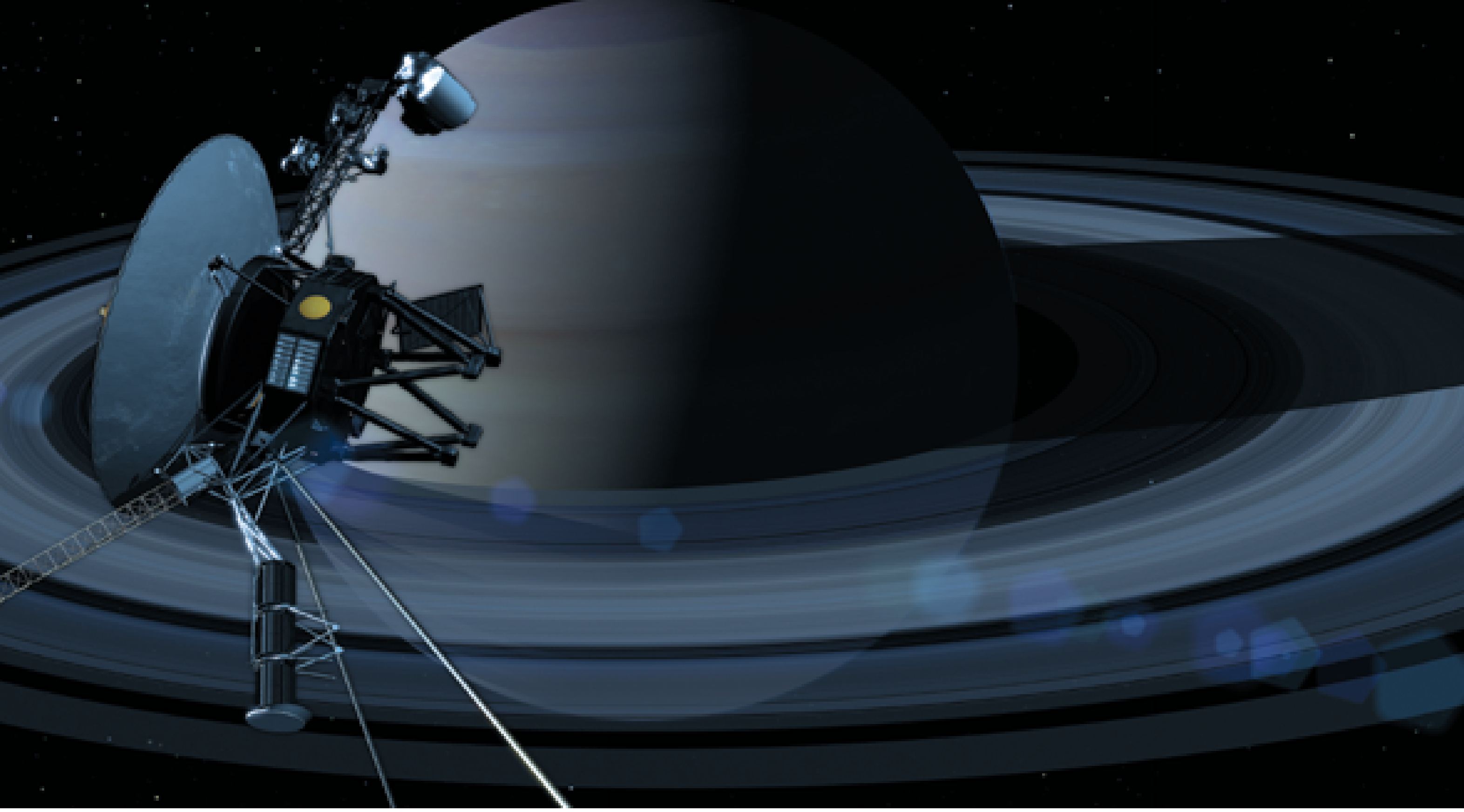
When the two *Voyager* spacecraft were launched in 1977, they followed a legacy of space exploration that was only two decades old, but which had accomplished much in that time. The first spacecraft to orbit Earth, *Sputnik 1,* was launched in 1957, and NASA's *Mariner 2* spacecraft passed Venus on December 14, 1962. *Mariner 4,* which swung by Mars on July 15, 1965—some 50 years ago—was the first to carry a camera. *Pioneer 10* and *11*'s flybys of Jupiter in the early 1970s, and *Pioneer 11*'s continuation to Saturn in 1979, gave a first glimpse of the complexity of these two planetary systems. The outer planets (Jupiter, Saturn, Uranus, and Neptune) as a group, however, remained basically unexplored until the two *Voyager* spacecraft arrived.

After the *Voyagers'* successful flybys of all four major outer planets, which they completed when *Voyager 2* departed Neptune in 1989, follow-on tasks could be taken on. The spacecraft began what has been dubbed the Voyager Interstellar Mission, with the objective of exiting the hot, extended atmosphere of the Sun, called the solar wind, and reaching the relatively cold gas in the local interstellar medium, the material filling the space between stars in a galaxy. Such a feat would give us our first direct experience of interstellar space. At the time of the *Voyagers'* launch, none knew how far this boundary would be, but estimates ranged from just beyond the orbit of Jupiter at 5 astronomical units (where one astronomical unit equals 150 million kilometers, the distance from the Sun to the Earth) to well beyond 50 astronomical units. As it turned out, the boundary was a lot farther than anyone had imagined.

As the solar wind travels out from the Sun, the plasma (electrons and ionized atoms) blows out a bubble that is confined by backward pressure from the galactic magnetic field; this region of influence of the solar wind is called the *heliosphere.* The galactic region of local interstellar space is dominated by plasma and cosmic rays originating from the explosions of supernovae over the last few million years. The surface where the radial flow of the solar wind brakes from about 300 kilometers per second to about 100 kilometers per second is called the *termination shock*. In this location, the solar plasma is deflected in azimuth and elevation, while its temperature increases from 10,000 to about 100,000 kelvin, creating what is called the *heliosheath* region. Eventually the radial plasma flow is expected to stop altogether; beyond that distance lies the *heliopause,* the border between the Solar System's plasma and magnetic field and those of the galaxy. Past this point, there is the possibility of a *bow shock,* a bit like the ripples of water created by a boulder in a stream, created by the Sun's movement, and the whole heliopause with it, through the interstellar medium. At the opposite side of the heliopause, there could form a cometlike tail of low-density plasma.

The *Voyager* spacecraft operated a full complement of instruments through all four planetary encounters, but for the interstellar mission only five are necessary and powered (*see the figure on page 286*). The others have been switched off to conserve the dwindling power supply from the probes' radioisotope thermal generators. The magnetometer, mounted at the end of a 13-meter boom, makes detailed measurements of the magnitude and direction of the ambient magnetic field. The plasma wave antennas measure the electric field components of plasma waves over the frequency range of 10 hertz to 56 kilohertz. The final three instruments are all mounted on the science boom: The cosmic ray sensor measures the intensity, composition, and spectra of high energy cosmic rays and relativistic electrons; the low energy

Stamatios M. Krimigis is principal investigator and Robert B. Decker is co-investigator for the Low Energy Charged Particle Experiment (LECP) on Voyager 1 *and* Voyager 2, *at the Johns Hopkins University Applied Physics Laboratory. E-mail for Krimigis: Tom.Krimigis@jhuapl.edu.*



Both *Voyager 1* and *Voyager 2* flew by Saturn, in 1980 and 1981, respectively, returning unprecedented images of that planet before continuing farther into space. *Voyager 1* became the first craft in interstellar space in 2012; *Voyager 2* will soon follow. (Image courtesy of NASA/JPL-Caltech.)

charged particle detector measures energetic particles at lower energies than the cosmic ray sensor, plus their flow direction; and the plasma sensor detects the properties of solar wind ions and electrons at energies as low as a few electron volts. (An electron volt is the energy an electron gains when it travels through an electrical potential difference of 1 volt; it's also equivalent to 1.6×10^{-19} joules.)

The last instrument is still functioning only on *Voyager 2* (although it's worth noting that after 38 years, that instrument has been the only one to fail). Each of the instruments performs a specific set of measurements, all of which are necessary to fully characterize the interplanetary and local interstellar medium.

A Complex Trajectory

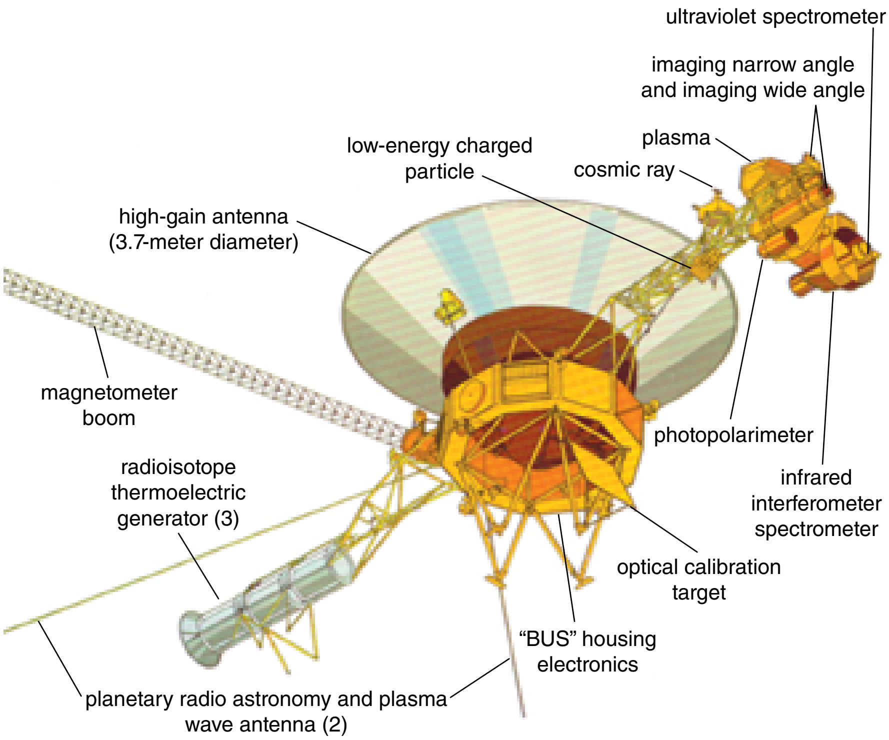
Getting the *Voyagers* on the path to their mission was no mean feat. The finite thrust of rockets taking off from Earth prevents them from propelling significant payloads much beyond the orbit of Jupiter, so a technique known as *gravity assist* is employed to accelerate spacecraft to higher speeds as well as change their direction. In this process, a well-planned flyby can use another planet's gravity to act like a slingshot, and add momentum to increase the energy of a spacecraft's orbit and propel it outward in the Solar System, much farther away from the Sun than its launch vehicle would have been capable of doing. A *retrograde* flyby, where the craft travels in the opposite direction of the planet's spin, can be used to subtract momentum and decrease energy, a move that can be used to change a craft's direction. If the flyby planet is at just the right part of the sky, then the gravity assist can send the spacecraft to another planet farther out. Gary Flandro of the Jet Propulsion Laboratory pointed out in 1965 that the outer planets lined up once every 175 years in such a way that a spacecraft could be propelled from one planet to the next in record time, and also surpass the Solar System *escape velocity* of about 7 kilometers per second, meaning that the Sun's gravity is no longer a factor on the spacecraft velocity. This opportunity would occur for a 1977 launch, whereupon detailed spacecraft and trajectory designs were undertaken at the Jet Propulsion Laboratory.

The science team decided early in the planning process that *Voyager 1*, which could have been targeted for a Pluto encounter after Saturn, should instead be directed to a close flyby of Titan, Saturn's largest moon; it would then follow a heliocentric trajectory toward the general direction of the *solar apex*, the direction of the Solar System's motion with respect to the local star group (*see the top figure on page 287*).

The *Voyager 2* trajectory was designed to take full advantage of the planetary alignment with encounters of not only Jupiter and Saturn, but also Uranus and Neptune. The spacecraft received a velocity boost with respect to the Sun of about 10 kilometers per second at Jupiter, 4 kilometers per second at Saturn, 2 kilometers per second at Uranus, and –3 kilometers per second at Neptune; the last one robbed some velocity because of a retrograde flyby needed to enable a close encounter with Neptune's moon Triton. However, *Voyager 2*'s heliocentric velocity of 15.6 kilometers per second still far exceeds the Solar System escape velocity.

The discoveries resulting from the planetary flybys of the *Voyagers* from 1979 through 1989 captivated the public imagination and demonstrated the excitement of robotic exploration like no other mission of the space era. The two spacecraft together revealed the complexity of the four planets, and unveiled the sheer beauty of worlds that had been only dimly perceived before.

Some of the findings of *Voyager* were totally unanticipated. For instance,



Only five of *Voyagers'* instruments are currently powered, and the plasma sensor is only functioning on *Voyager* 2. One central instrument, the low-energy charged particle detector, measures ions and electrons with energies from a few thousand to several tens of millions of electron volts. It uses detectors ranging in thickness from 2 micrometers to 2.5 millimeters; thinner detectors are required to identify the lowest energy particles. The most risky addition to the detector was a motor that rotated the fields of view of all detectors through a full 360 degrees. This moving part was considered very likely to fail, but the ones on each *Voyager* are still functioning after 38 years. (Image courtesy of NASA/JPL-Caltech.)

their images of Jupiter's moon Io showed lavas deposited via incessant volcanism, the first discovery of volcanism outside of our own planet. The power of Io's largest volcano is at least tenfold that of any on Earth. Another Jovian moon, Europa, was found to possess a surface of solid ice crisscrossed by cracks many tens of kilometers long, with colors suggesting upwelling from an underlying liquid ocean. Ganymede was shown to have an icy surface with both old terrain, as well as younger features that indicate later tectonic activity. And Callisto was characterized by long crater chains and a very old surface.

The *Voyagers* also imaged the Great Red Spot on Jupiter, first seen by Galileo Galilei in the 17th century, a persistent hurricane-like feature nearly 2.5 times the size of Earth. *Voyagers'* measurements also found that the outer planets have peculiar magnetic fields—as manifested in the tilt of their dipole axes with respect to their rotation axes, and their offsets from the center of their planets. For instance, Neptune has a tilt of 47 degrees and an offset of 55 percent. Such large values cannot be explained by current theories for the generation of planetary magnetism and have yet to be understood.

***Voyagers'* measurements found that the outer planets have peculiar magnetic fields, which have yet to be understood.**

Particle Plethora

Although their images of planets are breathtaking, some of the *Voyagers'* most enlightening data come from their measurements of the very small. During their interstellar mission, *Voyager 1* and *2* have measured distributions of energetic electrons and ions. These measurements include, for example, low-energy ions, mainly protons, with energies of 0.14 to 0.22 million electron volts, which originate in the heliosphere (*as shown in the upper right figure on page 288*) and galactic cosmic ray protons with energies of more than 70 million electron volts. The low-energy ions have speeds of 3.0 to 3.7 astronomical units per day, which, although relatively slow compared to that of the galactic cosmic rays (at 67 to 172.8 astronomical units per day), are still much faster than the mean solar wind speed of 0.25 astronomical units per day. Thus, the low-energy ions are free to propagate over large distances along the spiral-shaped magnetic field of the solar wind, enabling them to move far from their source regions.

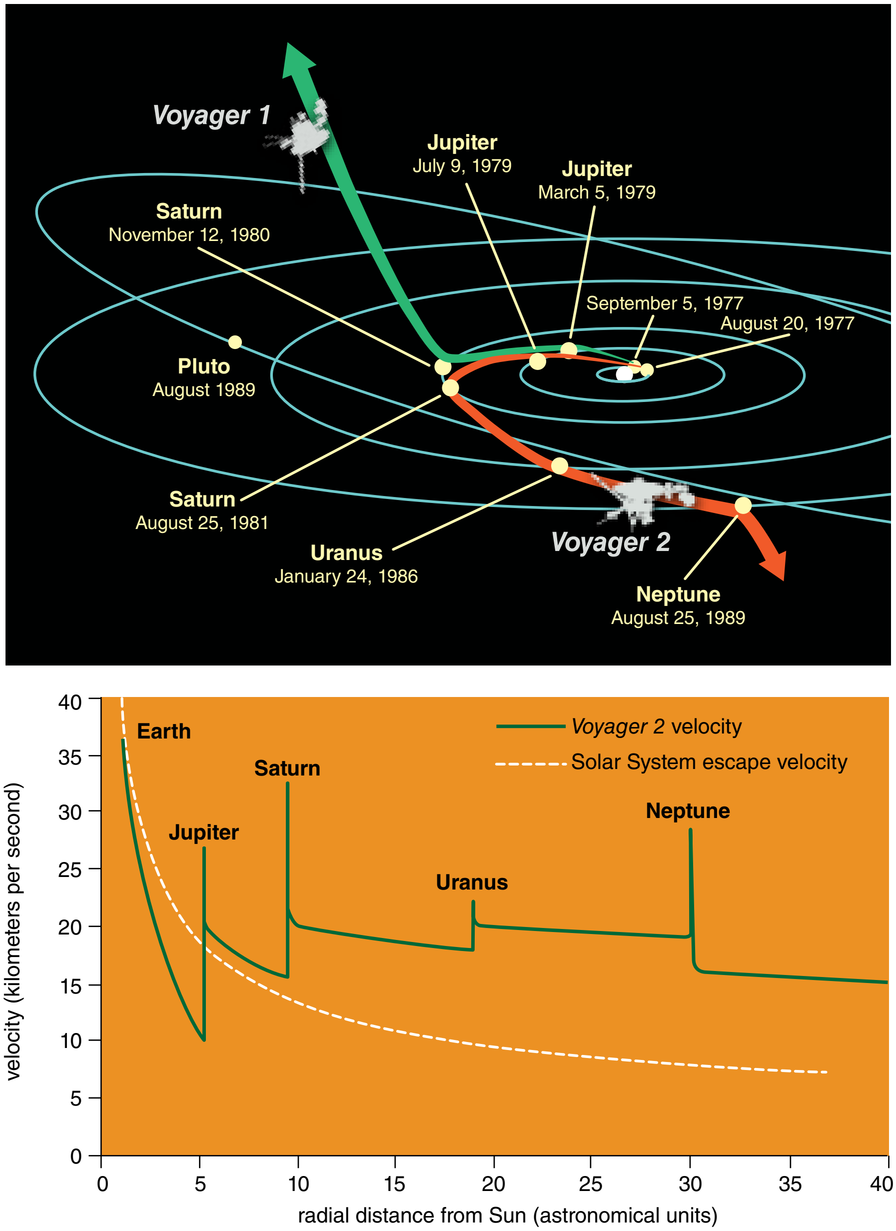
Voyager 1 and *2* spent their first 27 and 30 years, respectively, measuring phenomena in the solar wind. Energetic ions and electrons observed in the inner heliosphere originate mainly at the Sun during solar active years, when sunspot numbers are high. Charged particles accelerated by solar flares and at shock waves at the front edge of coronal mass ejections by the active Sun are called *solar energetic particles.* Charged particles accelerated at shock waves bounding the front and rear edges of solar wind stream interaction regions, called *corotating interaction regions*, form beyond a few astronomical units of the Sun during solar inactive periods. They tend to recur over several solar rotations with a period of about 26 days, as these regions rotate about the Sun like pinwheels.

Voyagers' measurements of the intensities of both populations of ions decreased roughly as the inverse square of the radial distance from the Sun, until about year 2000. Another ion source known as *pickup ions,* which play a major role beyond a few tens of astronomical units, particularly at the termination shock and in the heliosheath, are interstellar neutral atoms (mostly hydrogen, helium, and oxygen) that become ionized as they drift into the heliosphere at about 25 kilometers per second. Protons from galactic cosmic rays originate mainly from sources outside the heliosphere, such as at shock waves formed when stars in our galaxy explode to become supernovae. Such particles that enter the heliosphere must fight against the outflowing solar wind, which carries a spiral magnetic field with superposed small-scale magnetic fluctuations and larger-scale disturbances from elevated levels of solar activity. Consequently, there is a steady increase in galactic cosmic ray intensity as the *Voyagers* move farther from the Sun toward the local interstellar medium, but with an 11-year solar-cycle variation because of the inverse correlation between sunspot number peaks and galactic cosmic ray peaks.

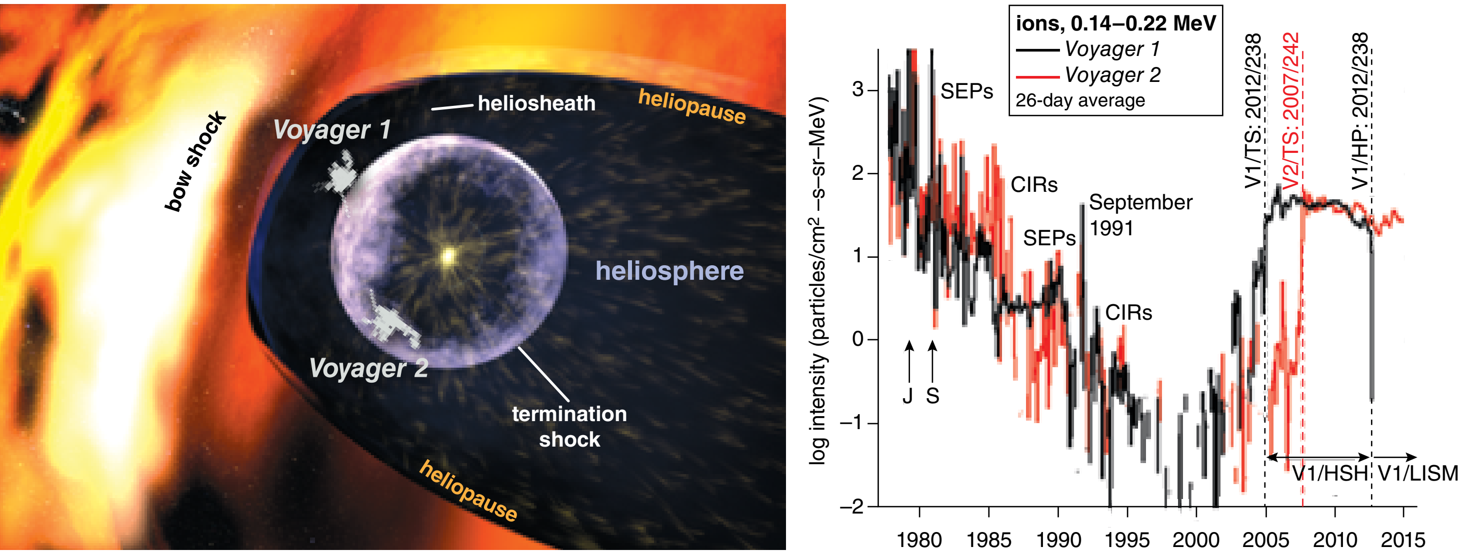
After the deep minimum of low-energy ion intensities reached in 1998–2000 (*as shown in the right figure on page 288*), ion intensities increased. *Voyager 1* crossed the termination shock on December 16, 2004, at 94 astronomical units, and explored the heliosheath (about 30 astronomical units wide along the *Voyager 1* path) for 8.2 years. We were warned of the impending termination shock crossing by precursor particles that had velocities highly collimated along the Sun's spiral magnetic field, but they arrived from a direction opposite to that predicted. The explanation was that the termination shock is blunted in its nose region by the asymmetric pressure imposed on the heliosphere by the tilted magnetic field of the local interstellar medium.

Voyager 2, following its very different path, crossed the termination shock several times between August 29 and 31, 2007, at 83.65 astronomical units from the Sun. The multiple crossings are believed to have resulted either from waves on the termination shock surface or from quasi-periodic modification of its internal structure by nonthermal pickup ions. *Voyager 2* is now in the heliosheath, about 22 astronomical units beyond the termination shock, and is heading toward the heliopause. The heliosheath is a relatively steady and, at least at the locations of the two *Voyagers*, uniform reservoir of low-energy ions with high intensities. The energies of protons measured in this reservoir extend from at least 0.03 to about 30 million electron volts.

The plasma sensor on *Voyager 2* enabled measurements of the thermal plasma before and after the craft crossed the termination shock. However, based on the plasma data alone, the calculated sound speed behind the shock was less than the flow speed in that region, which implied that the termination shock was in fact not a shock in the technical sense that the pre-shock flow should be decelerated by the shock to a subsonic speed. *Voyager* investigators soon realized that the pressure of the nonthermal ions must be included in calculating the sound speed in the heliosheath, but these particles are well below the energy range accessible to the low energy charged particle instrument and the intensity measurable by the *Voyager* plasma sensor. We know that this distribution of shocked and heated pickup protons must exist in the heliosheath because it is this population of hot protons that undergo charge exchange with inflowing cold, neutral hydrogen atoms, and thus provide the energetic neutral atoms recently observed remotely by two other spacecraft, the *Interstellar Boundary Explorer* (IBEX) located at



The trajectories of *Voyager 1* and *2* (*top*) during their planetary encounters show that *Voyager 1* was sent northward of the ecliptic at Saturn in 1980, and *Voyager 2* southward at Neptune in 1989. The crafts could not have traveled so far without a gravity assist from each planet visited. The bottom graph compares the speed needed to escape the Sun's gravitational pull (*dashed white line*) and *Voyager 2*'s speed (*green line*) versus radial distance from the Sun. The net effect of gravity-assisted speed gains from Jupiter, Saturn, and Uranus enabled *Voyager 2* to exceed the velocity needed to escape the Solar System. (Image courtesy of NASA/JPL-Caltech.)



A view of the Sun (*at center of circle*) and the heliosphere, the bubble created by the interaction between the outflowing solar wind and the back pressure on it from the interstellar medium. The solar wind is decelerated from supersonic speeds at the termination shock. The flow in the heliosheath is diverted so it flows down the long tail of the heliosphere. The graph at right shows how charged particles from the Sun gradually decreased until the *Voyagers* crossed the termination shock. (Left illustration courtesy of NASA/Walt Feimer; right graph courtesy of the authors.)

1 astronomical unit and *Cassini* at 10 astronomical units.

A remarkable finding from *Voyager 1*'s passage through the heliosheath was the unexpected evolution of the plasma flow velocity there. As shown in the figure on page 289, the radial component (labeled R) of the solar wind velocity is slowed and deflected across the termination shock, and expected in the heliosheath to acquire meridional (N) and azimuthal (T) components. The plasma detector on *Voyager 1* has not functioned since 1981. However, the low-energy charged particle instrument measures low-energy ions in eight 45-degree sectors that cover a full circle lying in the R–T plane. In situations when the low-energy ions tended to move with the plasma, such as in the heliosheath, these directional data were used in lieu of plasma detector data to estimate the R-component and T-component of plasma flow in the heliosheath.

A long-running debate on whether or not *Voyager 1* was in interstellar space became a dispute between observations versus theory and modeling.

As *Voyager 1* moved deeper into the heliosheath, the T-component remained small and relatively constant at about –20 to –40 kilometers per second, whereas the R-component decreased from a peak of 100 kilometers per second at 97 astronomical units, to 0 kilometers per second at 113 astronomical units. This change in itself would not be a surprise, because the radial velocity was expected to vanish just before the heliopause crossing to allow the heliosphere plasma to flow parallel to the heliopause surface. However, *Voyager 1* measured this zero radial velocity for an additional 2 astronomical units and then entered a region from 115 to 121 astronomical units in which the velocity fluctuated, but was on average –15 kilometers per second. In other words, it is directed sunward, suggesting that the heliosheath plasma was somehow coupled to the inflowing local interstellar medium plasma.

We all wondered: What had happened to the heliosheath flow? Was radial flow being diverted into meridional flow?

To help answer this question, in early 2011 ground controllers instructed *Voyager 1* to roll about the Earth-spacecraft line through about 90 degrees, and hold this attitude for several hours every few months to enable estimates of the meridional velocity. These rolls continue as *Voyager 1* moves through the local interstellar medium. So far, the results show that meridional velocity was at most a few kilometers per second in 2011 and 2012, meaning that radial flow had not been diverted into the meridional direction. Several models have been advanced to explain the evolution of the plasma flow in this so-called stagnation region extending from 113 to 121 astronomical units. There is as yet no consensus that favors any one model over the others.

Solar Exit

Despite the surprises encountered at the edge of the heliosheath, our expectation was that once the heliopause was crossed, the disappearance of solar wind plasma would be simply accompanied by an increase in galactic cosmic rays and a turn in the magnetic field direction to that estimated for the galactic field. The reality revealed by the *Voyager* measurements turned out to be quite different.

The galactic cosmic ray intensity increased in steps, first on May 7, 2012, then up and down starting on July 28, again on August 9, and finally on August 25 at a distance of 121.6 astronomical units from the Sun. The last increase was most pronounced along the direction of the field, but less so perpendicular to it. The expectation from theory was that galactic cosmic rays would not have a preferential direction in the local interstellar medium. Coincident decreases in solar particles also were not evenly distributed in direction.

In contrast to the galactic cosmic ray increases, particles escaped quickly along the field but much more slowly perpendicular to it. The magnetic field increased in magnitude coincident with the particle changes, but the direction stayed nearly constant during the short events and continuously so after the final discontinuity on August 25, 2012. Furthermore, the magnetic field magnitude reached a value more than four times that seen throughout the heliosphere.

The azimuthal direction of the solar magnetic field inside the cavity formed by the solar wind was expected to change to that of the north-south interstellar field, but did not. And the galactic cosmic rays have been anisotropic (not uniformly distributed), except for some intervals over the past year, and continue to exhibit this behavior to this day. These confusing details prompted a long-running debate on whether *Voyager 1* was or was not in interstellar space that became a dispute between observations versus theory and modeling.

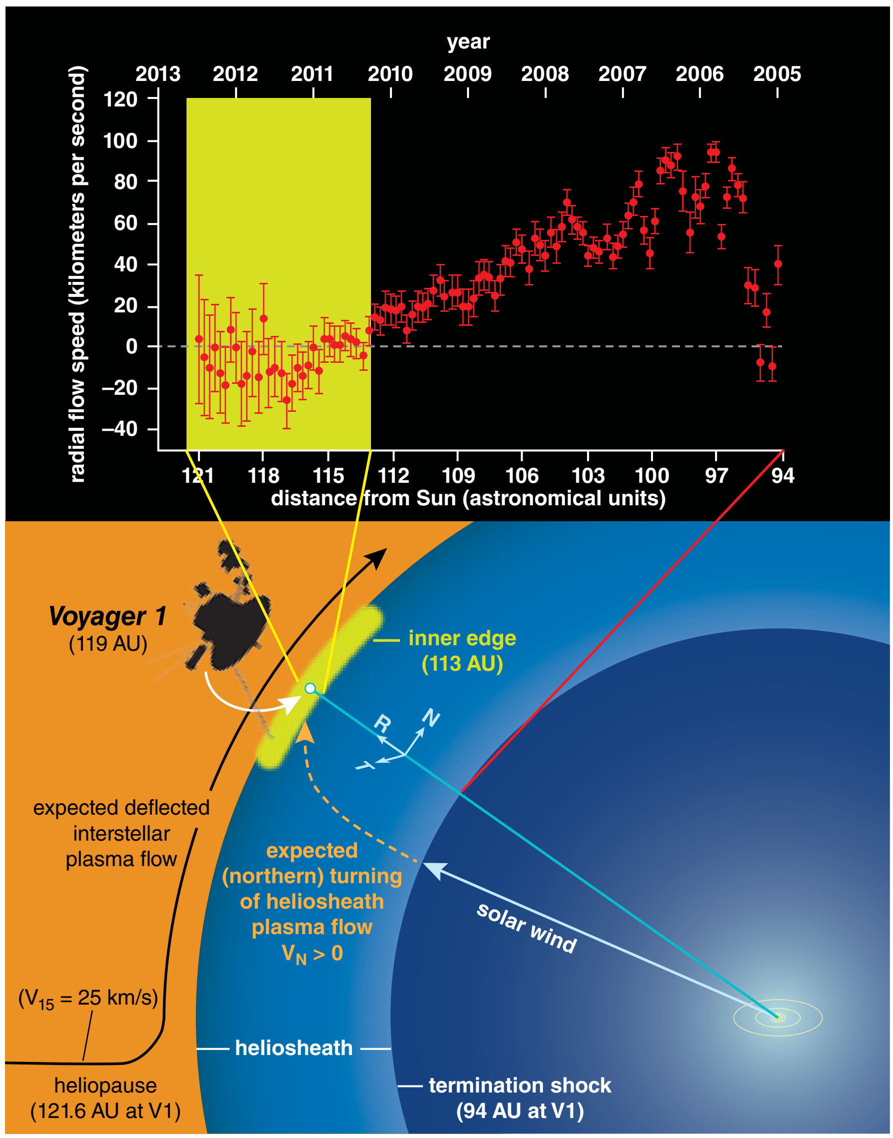
On the one hand, the galactic cosmic rays increased as expected, but were anisotropic and their distribution was ordered by the magnetic field (both unexpected events). The solar material disappeared, as expected, but its escape was also ordered by the magnetic field, which was unexpected. And the magnetic field direction did not change to that estimated for the local interstellar medium—indeed, it remained nearly the same as that in the solar wind cavity, leading some to suggest that the boundary crossed on August 25, 2012, was not the real heliopause.

Initially it was assumed that the lack of directional change in the magnetic field following the heliopause boundary crossing could be taken as a proxy that there was still solar plasma present at the spacecraft's location after August 25 (which was not possible for *Voyager 1* to measure because of its inoperative plasma sensor). The counter to that argument was that solar material down to energies of 40,000 electron volts was undetectable, and at somewhat higher energies, the quantity was down by more than 10,000.

In fact, a variety of arguments, also based on energetic neutral atom images from the *Cassini* spacecraft in orbit around Saturn (a new technique that can image the entire sky remotely), suggested that solar material had probably disappeared at any energy. Those arguments led to an estimate for the location of the heliopause at 121 astronomical units, as observed. Then there was the persisting anisotropy of galactic cosmic rays that was not predicted by any theoretical model.

The missing measurements necessary to resolve the conundrum were plasma density, temperature, and direction, which the plasma detector would have provided. In April 2013, the plasma wave antennas, not having detected any wave activity for the past several years, began to see electron plasma frequency oscillations. These measurements can be converted to density, and that was found to be close to the estimated value expected for the relatively cold galactic plasma. By contrast, the density in the heliosphere, measured by *Voyager 2* with its still-operating plasma detector, is about 50 times smaller.

Although the plasma temperature cannot be measured, the high density



The radially (R) flowing supersonic solar wind is slowed across the termination shock and was expected to acquire meridian (N) and azimuthal (T) components in the heliosheath. Measurements of plasma flow (*top*) show a stagnation region (*yellow*) in which the radial velocity at *Voyager 1* went from zero to highly fluctuating negative values before crossing the heliopause. (Images courtesy of NASA/JPL-Caltech and JHU/APL, modified by the authors.)

is sufficient to prove that *Voyager 1* has moved beyond the heliopause and is now indeed moving through the local interstellar medium. It must be noted that the sudden appearance of electron plasma oscillation was most likely due to solar activity back in March 2012, and the resulting plasma cloud arrived at *Voyager 1* some 13 months later. In fact, a more detailed examination of the plasma wave antenna data revealed that similar, but lower, density oscillations were seen in the October–November time frame, most likely originating from earlier solar activity, that extrapolate into a crossing of the heliopause in late August 2012.

The galactic environment has not been the calm and benign regime that we all had expected.

On September 9, 2013, we hosted a meeting at Johns Hopkins University of the entire *Voyager* team, at which the totality of the *Voyager 1* observations was presented, analyzed, and contrasted with prevailing models. There was general agreement that *Voyager 1* was, indeed, in the local interstellar medium and that the heliopause crossing took place on August 25, 2012, at a distance of 121.6 astronomical units (with light from the Sun taking 16 hours and 54 minutes to reach this distance), some 35 years after the spacecraft was launched from Cape Canaveral.

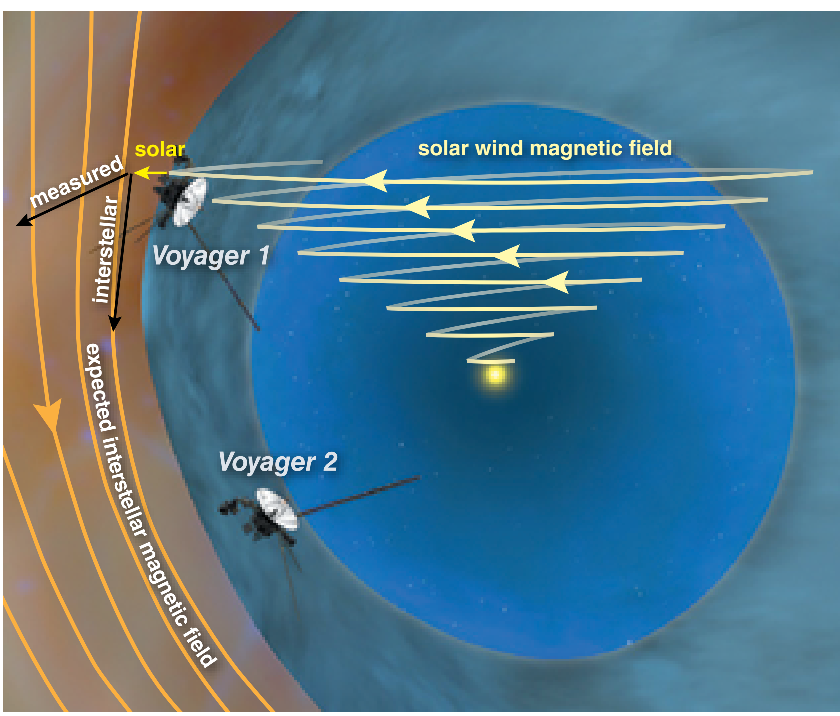
After entering interstellar space, *Voyager* has continued to deliver surprises. The galactic environment has not been the calm and benign regime that we all had expected. Data on galactic cosmic ray intensities through the end of 2014 show that there exist periods of quiet isotropy, followed by incidents of anisotropy, as if an occasional "tsunami" of activity perturbs the upstream medium after crossing the heliopause. These pressure waves most likely originate at the Sun, propagate through the heliosphere, and eventually arrive at the location of *Voyager 1*.

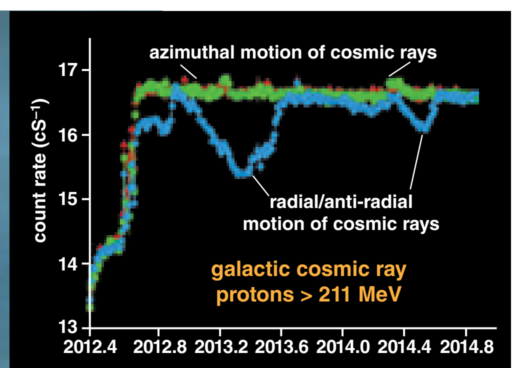
There is apparently a region beyond the heliopause, perhaps leading to a bow shock, beyond which the solar influence wanes. If so, that distance is at least 10 astronomical units away from the heliopause, where *Voyager 1* is currently located (131 astronomical units from the Sun).

Teamwork

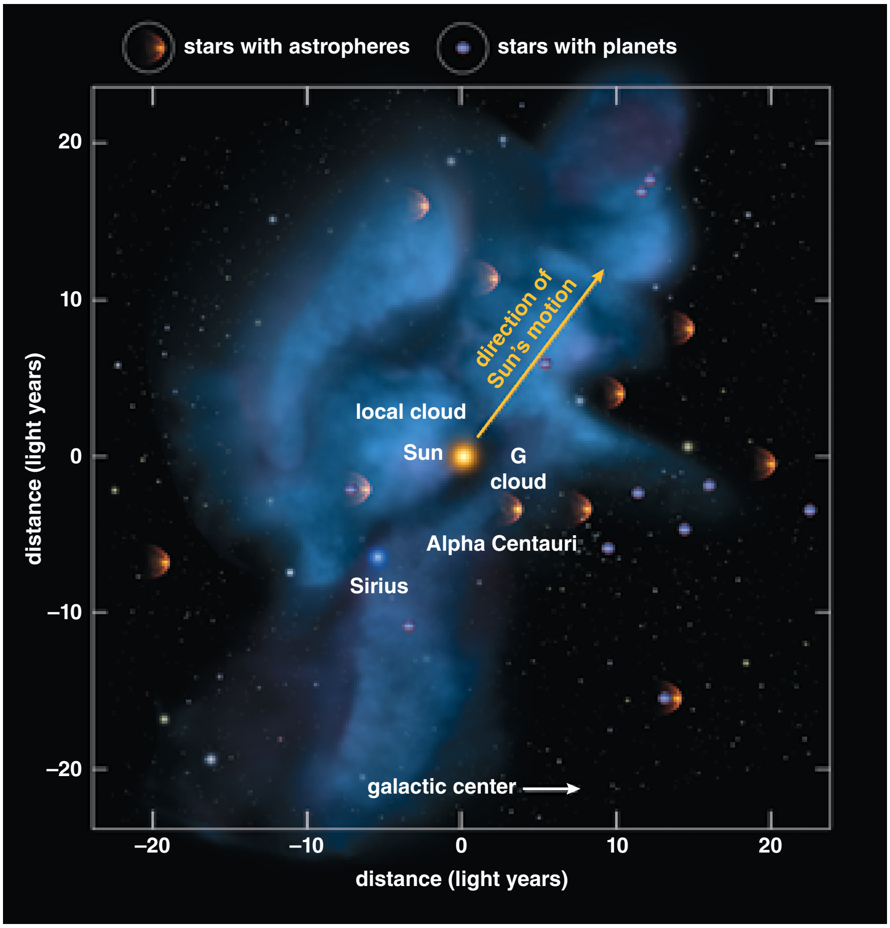
The odyssey of the two *Voyagers*, launched in 1977 and now on their 38th year in space, represents the epic mission of the space era. *Voyager 1* has accomplished far more than its originators had any right to expect, and its journey of discovery has yet to end. The anticipated benign environment beyond the heliopause is anything but that so far, and new phenomena are revealed as it moves farther away. *Voyager 2*, at 108 astronomical units and 31 degrees south of the ecliptic, is currently exploring the heliosheath and there is no clear indication when it may cross the heliopause into the upstream region. It has the advantage of a working plasma instrument that will resolve some of the science questions raised by the *Voyager 1* crossing, such as the temperature and directional flow of the galactic plasma. Also, *Voyager 2* could conceivably answer the question of the existence of a bow shock, because *Voyager 1* may not provide an unambiguous answer in the absence of plasma data.

The question comes up often about how long the *Voyagers* will last. The limiting factor is their radioisotope thermoelectric generators that convert heat from the radioactive decay of plutonium-238 into electrical current to operate the spacecraft and instru-





The solar wind spiral magnetic field (*pale yellow lines*) were expected to move perpendicularly past the heliopause (*dark yellow lines*) but when measured (*black arrow*) the field remained very similar to the direction of the heliospheric one. Measurements of galactic cosmic ray protons (*above*) also showed that they were unexpectedly unevenly distributed with respect to their arrival directions at *Voyager 1*. (Images from NASA/JPL-Caltech and JHU/APL, modified by the authors.)



The motion of the Solar System relative to the local interstellar medium is inferred from astronomical observations. The image also shows the known stars with astropheres—equivalent to the Sun's heliosphere. *Voyagers'* travels past the heliosphere could tell us more about the makeup of our wider region of space. (Image courtesy of NASA/Goddard/Adler/U. Chicago/Wesleyan.)

ments. The half-life of plutonium-238 is 87.7 years, so the initial power at launch of 465 watts is now down to about 264 watts. In its current configuration, it is likely that *Voyager 1* can fully power all its instruments through about 2020, at which time there will commence power-cycling of instruments. By about 2025, there probably will not be enough power to operate even one instrument, at which time *Voyager* operations are expected to cease. Although disappointing, it will be difficult to complain about the performance of these marvelous spacecraft, as the initial warranty of four years expired long ago.

It is often pointed out that *Voyager* did not leave the Solar System per se, but only the heliosphere, the ionized atmosphere of the Sun. Put another way, they escaped the region of the Sun's chemical environment but not its gravitational environment. This distinction is strictly true, because the comets in the Oort cloud are loosely bound into solar orbit at distances up to 100,000 astronomical units. Note that Alpha Centauri, the nearest star to our own, is at about 280,000 astronomical units. At its present speed of 17 kilometers per second, *Voyager 1* will pass a star named AC+79 3888, located in the constellation of Camelopardalis, in about 40,000 years. Even that will not be a very close encounter in *Voyager*'s long, lonely path.

It is unfortunate that none of the world's space agencies are currently planning an interstellar probe as a follow-on to *Voyager*, even though the technology exists to attain at least 200 astronomical units in 25 years, and even 500 astronomical units in 50 years. Thus, knowledge gained by the *Voyager* mission is unlikely to be duplicated and advanced for at least another generation. We owe a tribute to the men and women of the latter part of the 20th century who built a wonderful set of spacecraft, and their current-day successors who have maintained and operated them for the benefit of human knowledge.

Bibliography

Burlaga, L. F., N. F. Ness, and E. C. Stone. 2013. Magnetic field observations as *Voyager 1* entered the heliosheath depletion region. *Science* 341:147–150.

Burlaga L. F., et al. 2008. Magnetic fields at the solar wind termination shock. *Nature* 454:75–77.

Decker, R. B., et al. 2008. Mediation of the solar wind termination shock by non-thermal ions. *Nature* 454:67–70.

Decker, R. B., et al. 2005. *Voyager 1* in the foreshock, termination shock, and heliosheath. *Science* 309:2020–2024.

Gurnett, D. A., W. S Kurth, L. F. Burlaga, and N. F. Ness. 2013. In situ observations of interstellar plasma with *Voyager 1*. *Science* 341:1489–1492.

Krimigis, S. M., et al. 2013. Search for the exit: *Voyager 1* at heliosphere's border with the galaxy. *Science* 341:144–147.

Krimigis, S. M., D. G. Mitchell, E. C. Roelof, K. C. Hsieh, and D. J. McComas. 2009. Imaging the interaction of the heliosphere with the interstellar medium from Saturn with *Cassini*. *Science* 326:971–973.

Krimigis, S. M., E. C. Roelof, R. B. Decker, and M. E. Hill. 2011. Zero outward flow velocity for plasma in a heliosheath transition layer. *Nature* 474:359–361.

McComas, D. J., et al. 2009. Global observations of the interstellar interaction from the *Interstellar Boundary Explorer* (IBEX). *Science* 326:959–962.

Richardson, J. D., J. C. Kasper, C. Wang, J. W. Belcher, and A. Lazarus. 2008. Cool heliosheath plasma and deceleration of the upstream solar wind at the termination shock. *Nature* 454:63–66.

Stone, E. C., et al. 2013. *Voyager 1* observes low-energy galactic cosmic rays in a region depleted of heliospheric ions. *Science* 341:150–153.

Zank, G.P. 2015. Faltering steps into the galaxy: The boundary regions of the heliosphere. *Annual Reviews of Astronomy and Astrophysics* 53, in press.

For relevant Web links, consult this issue of *American Scientist Online:*

http://www.americanscientist.org/issues/id.115/past.aspx