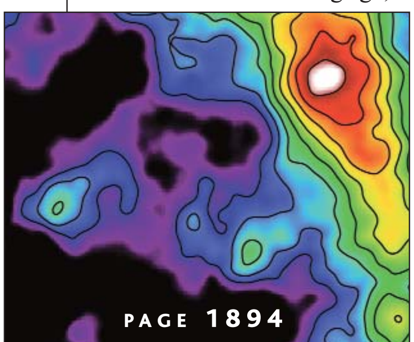
INTRODUCTION

Welcome to the Dark Side: Delighted to See You

Dark stars, the dark age, dark matter, and dark energy are the major components of the dark side of the universe: 96% of the universe consists of mass and energy we can't see and don't really understand. Fortunately, the badly outnumbered 4% of luminous matter feels the dark side through gravity and other forces. Stellar struggles with the dark side, which we can see through gravity and electromagnetic emissions, have much to tell us about the bulk of the universe. Here five intrepid astronomers and two News writers review what we know or think we know about these epic battles throughout cosmic time.

Perhaps the best-understood component of the dark side is the dark stars called black holes. Begelman (p. 1898) explains that black holes are common. By studying black holes in the center of the Milky Way and other galaxies, astronomers have discovered that their masses are correlated with certain types and masses of galaxies, suggesting that either the black hole knows about the structure of the stars in the galaxy or the stars know about the black hole through other indirect forces. The answer to this "who-saw-who-first" question may hold the key to explaining how black holes and galaxies form.

Long ago, the universe was dark and there were no stars. Miralda-Escudé (p. 1904) reviews what we know about this dark age. He concentrates on the hints of light at either end of the dark age: the cosmic microwave background radiation that dispersed right after the big bang, at a redshift of 1100, and the first stars that formed about 75 million years later, at a redshift of about 38. Although we have not seen the first population of stars, we can observe stars as far back as a redshift of about 6. During the dark age, dark matter clumped together, creating density fluctuations that could collapse and form stars.



PAGE 1894

The first stars formed from this dark matter, which provided the blueprint, the DNA, for cosmic structure and its rate of evolution. Ostriker and Steinhardt (p. 1909) discuss the possible types of dark matter (it is not a double helix) that may now account for about 26% of the universe. The "cold dark matter" model says that dark matter is made up of cold particles such as neutralinos or other weakly interacting massive particles (WIMPs). More sinister-sounding varieties, such as self-annihilating or repulsive dark matter, may also exist and battle against WIMPs to determine the fate of ordinary matter. Irion's News article (p. 1894) describes how researchers plan to map the distribution of dark matter throughout the universe by analyzing its subtle effects on the light from distant galaxies.

Nowadays, about 70% of the universe is dominated by dark energy, which is the dark-side component we understand the least. Evidence for dark energy comes from hundreds of type Ia supernovae, detected as far back as a redshift of 1.8. As Kirshner explains (p. 1914), supernovae show that the expansion of the universe has been accelerating over the past 7 billion years, and the acceleration is caused by dark energy. By extending observations of supernovae further back in time, we should be able to see when the universe shifted gears from deceleration caused by clumpy, gravitationally attractive dark matter to acceleration caused by less clumpy, gravitationally repulsive dark energy. As Seife reports (p. 1896), cosmologists hope that this cosmic tipping point, along with a better understanding of the physical properties of dark energy, will provide beachheads for future forays into this murkiest province of dark-side science.

Science Online sheds further light on darkness with links to missions, experiments, and papers bearing on various topics in cosmology (www.sciencemag.org/feature/data/darkside/).

May the dark energy be with you as we struggle to understand the darkness of space and time.

–LINDA ROWAN AND ROBERT COONTZ

CONTENTS

NEWS

REVIEWS

Science

CREDIT: J. A. TYSON AND D. M. WITTMAN/DEEP LENS SURVEY

NEWS

The Warped Side of Dark Matter

Weak gravitational lensing, a subtle distortion of all distant galaxies, promises the most direct way of mapping the universe we can't see

Imagine flying over a mountain range on a moonless night. You know that peaks loom below, but you can't see them. Suddenly, specks of light pop into view: isolated country homes, dotting the hilly slopes. The lights outline part of the massive edifice, but your mind grasps that the darkness hides something far larger.

Astronomers face a similar situation. In recent years, their research has confirmed that the luminous universe—our sun, our galaxy, and everything that shines—makes up but a wee bit of all there is. Instead, the strange new recipe calls for more than one-quarter "dark matter" and two-thirds "dark energy." This is the universe your teacher never told you about: matter of a completely unknown nature and energy that hastens the expansion of the cosmos toward future oblivion.

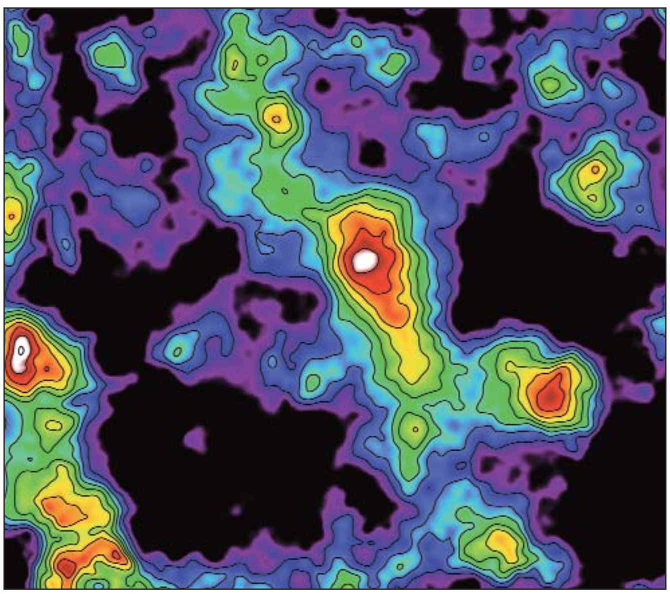
To divine the properties of dark matter, astronomers first must find out where it is. And to learn how dark energy controls the fate and shape of the universe—including how matter is distributed—they must trace how the dark matter clumped together over time. But they can't see it; all they have are some bright dots in a vast, mountainous wilderness.

That's about to change. Researchers are refining an exciting new technique that relies on the warping of space itself to reveal dark matter. Called weak gravitational lensing, the method exposes dark matter by tracing the subtle distortions it imparts to the shapes and alignments of millions of distant galaxies. The effect isn't obvious to the eye, yet it alters the appearance of every remote galaxy. Although widespread detection of this "cosmic shear" first hit journals just 3 years ago, several teams worldwide have embarked on major new surveys in a race to exploit its potential. Indeed, astronomers now feel that weak lensing will become a cornerstone of modern cosmology, along with studies of the cosmic microwave background radiation and distant explosions of supernovas.

"I no longer regard galaxies as tracers of the cosmos," says astronomer Richard Ellis of the California Institute of Technology (Caltech) in Pasadena. "We now have the confidence to go after the real physics. Let's image the dark matter directly; we have the tools to do it. Weak lensing is one of the cleanest cosmic probes of all."

Line up and stretch

Weak lensing is akin to the far more spectacular process called strong gravitational lensing. In the latter, the intense gravity of galaxies or clusters of galaxies bends and magnifies light from more distant objects as the light travels toward Earth. Strong lensing can split a single quasar into four images or distort remote clusters into dizzying swirls of eerie arcs. These funhouse mirrors in space, captured exquisitely by the Hubble Space Telescope, are vivid displays of the pervasive light-bending effects in Albert Einstein's general theory of relativity.



Brought to light. Weak gravitational lensing exposed these patches of dark matter, otherwise hidden from telescopes.

CREDIT: J. A. TYSON AND D. M. WITTMAN/DEEP LENS SURVEY

Relativity also causes weak lensing, but without such drama. "Strong lensing is like pornography: You know it when you see it," says astronomer R. Michael Jarvis of the University of Pennsylvania in Philadelphia. "Weak lensing is like art." And like art critics, astronomers have honed their perception to see weak lensing where others see a featureless array of galaxies.

The array is a background of millions of faint blue galaxies, first recognized in the late 1980s. This "giant tapestry," in the words of astronomer Ludovic Van Waerbeke of the Institute of Astrophysics in Paris (IAP), freckles any exposure of the heavens by research telescopes with mirrors larger than 2 meters across. The galaxies date to a time when the universe was less than half its current age, and they are everywhere astronomers look.

Although each galaxy looks like a disk or an elongated blob, the mathematical average of a large number of them is a round shape. In a similar way, the galaxies should not line up in a special direction; on average, their orientations should be random. Weak lensing, induced by the tugging of dark matter between us and the faint galaxies, leaves patterns in those shapes and alignments at a tiny level of distortion: about 1%. Finding the patterns thus becomes a statistical game. "Each galaxy is like a little stick on the sky, and we want to measure its elongation and orientation," Van Waerbeke says. To see that signal reliably, astronomers must take steady photos of the galactic tapestry. Useful images typically capture at least 20,000 galaxies in a patch of sky the size of the full moon—one-fifth of a square degree.

Then, using the physics of relativity, the researchers convert the slight distortions into a plot of all of the mass—both luminous and dark—along the path between Earth and the distant galaxies. This plot (see figure at left) is a two-dimensional projection; it doesn't reveal the distance to each blob. Even so, it exposes unseen mountains of mass whose gravity changes the appearance of everything on their far sides. "To see this, we don't have to make assumptions about what the dark matter is," says astronomer Jason Rhodes of Caltech. "It's the most direct way to simply measure everything that's there."

Of course, there are complications. The atmosphere blurs galaxies, telescopes jitter, and electronic detectors have flaws. Statistics quickly degrade unless images are rock solid over a wide patch of sky. But the promise of weak lensing was so potent in the late 1990s that a spirited race pushed astronomers to tackle these technology issues.

When success came, it came with a flash: four nearly simultaneous papers in March 2000 from groups in Canada, Europe, and the United States on the first detections of cosmic shear over large areas.

Since then, teams have extended their efforts in two ways. Some look at broader sweeps of the sky with modest telescopes, such as the 3.6-meter Canada-France-Hawaii Telescope (CFHT) on Mauna Kea, Hawaii, and the 4.2-meter William Herschel Telescope on La Palma, Canary Islands. Those projects aim to examine as many dark-matter patches as possible in a sort of population survey, improving the overall statistics of their distribution through the universe. Others use big telescopes, including one of the European Southern Observatory's four 8.2-meter Very Large Telescopes on Cerro Paranal, Chile, and one of the twin 10-meter Keck Telescopes on Mauna Kea, to zero in on a few distant regions with greater depth.

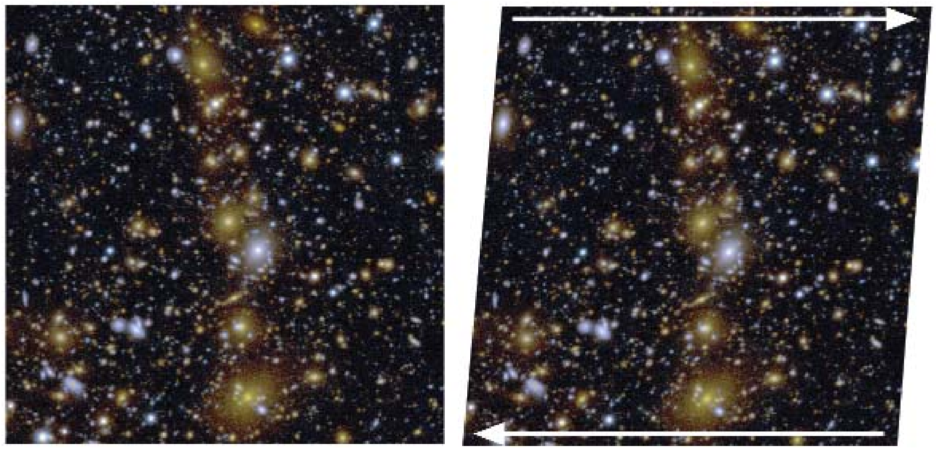
Most of the invisible mass found by weak lensing is mingled with ordinary galaxies visible in either optical light or x-rays. However, some teams claim to have spotted concentrations of matter with no associated galaxies at all. These truly dark clusters, if they are real, would betray the universe's dirty secret: Big piles of mass don't necessarily come with lights attached.

Most agree that shaky statistics make those claims vague for now, but the fundamental lesson is valid. "The ratio between emitted light and underlying mass changes quite considerably" from cluster to cluster, says theorist Matthias Bartelmann of the Max Planck Institute for Astrophysics in Garching, Germany. "This is something unexpected."

The implication is profound. Astronomers cannot rely on large-scale surveys of galaxies alone to trace the history of how matter has assembled in the universe. But that history is critical to unraveling the riddle of dark energy. As Bartelmann notes, dark energy apparently has exerted its greatest influence during the past several billion years. As the expansion of space carried matter farther apart, gravity became less effective at slowing the expansion. Meanwhile, dark energy—manifested as a self-repulsion within the fabric of space itself—grew dominant (see p. 1896).

Theorists are eager for an atlas of how dark matter clumped together to help them see what makes dark energy tick. "We have no other way to calibrate how structures formed in an unbiased way in the last one-third of cosmic evolution," when dark energy's sway took hold, Bartelmann says. "Weak lensing is without competition in that field."

Teams already are taking a first crack at measuring the clumpiness of dark matter. In essence, a smooth spread of dark matter between us and a distant galaxy has a minor lensing effect, whereas blobs of the stuff enhance the weak-lensing signal—just as marbled glass on a thick shower door distorts light more than plate glass does. Even with current statistics, results from weak-lensing surveys help pin down numbers for the mass content and expansion rate of the universe, according to a paper in press at *Physical Review Letters* by astrophysicist Carlo Contaldi of the Canadian Institute for Theoretical Astrophysics in Toronto and colleagues. "The combination of [cosmic microwave background radiation] and weak-lensing data provides some of the most powerful constraints available in cosmology today," the team writes.



Shear science. Distant galaxies show random shapes and orientations (*left*) unless intervening dark matter shears those patterns in a subtle but detectable way (*right*).

CREDIT: ADAPTED FROM DAVID WITTMAN/BELL LABS/LUCENT

Another promising way to chart dark matter's behavior is "3D mass tomography," named by a pioneer of weak lensing, astrophysicist J. Anthony Tyson of Lucent Technologies' Bell Laboratories in Murray Hill, New Jersey, and his colleague David Wittman. Researchers can gauge the distances to blobs of dark matter by crudely estimating the distance to each distorted galaxy in the background tapestry. Light from the most distant galaxies crosses the greatest chasm of space and gets lensed most severely, whereas relatively nearby galaxies aren't affected as much.

By correlating the distortions of galaxies with their rough distances, Tyson's team can convert the 2D projections of total mass into 3D volumes. That reveals where the dark-matter mountains are in space with 10% to 20% accuracy. Using data from the 4-meter National Optical Astronomy Observatory telescopes at Kitt Peak, Arizona, and Cerro Tololo, Chile, the group has derived locations for about two dozen dark clusters. When the astronomers complete their survey of 28 square degrees of the sky in 2004, they expect to identify 200 clusters out to a distance of about 7 billion light-years, says Wittman.

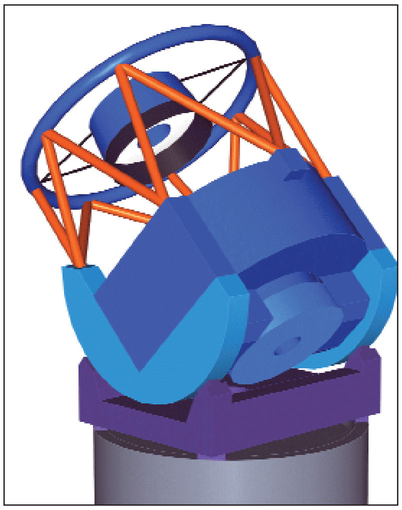
Take a wider view

Still, Tyson's program and all other efforts face similar problems: Images aren't sharp enough, deep enough, or wide enough. "The facilities we have worldwide don't yet have the light grasp and field of view required to get the scientific promise out of weak lensing," Tyson says.

Astronomers are launching a second generation of cosmic-shear surveys that should achieve some of that promise. Foremost is the CFHT Legacy Survey, powered by the biggest astronomical camera ever built: MegaPrime, which can take sharp images of a full square degree of sky (five full moons). The 170-square-degree survey, set to begin within weeks, will consume 100 nights per year for 5 years on the CFHT. Goals include searching for supernovas and nearby transient objects, such as hazardous asteroids. However, the weak-lensing part of the survey—led by IAP astronomer Yannick Mellier—has the community abuzz. "MegaPrime is a magnificent instrument, and this survey will be a landmark in the field," says Caltech's Ellis.

A hot competitor is one of CFHT's neighbors under the crisp Mauna Kea skies: Japan's 8.2-meter Subaru Telescope and its new Suprime-Cam. Although its field of view is just one-fourth that of MegaPrime, Suprime-Cam has won equal raves for its image quality. Moreover, Subaru's mirror has more than four times as much light-collecting power as does CFHT. That will let the Japanese team examine lenses in far greater detail. The astronomers plan to use 3D tomography to pinpoint the masses, distances, and rough shapes of hundreds of dark entities. "We would like to publish the first mass-selected object catalog [of dark-matter lenses] in a timely manner," says team leader Satoshi Miyazaki of the National Astronomical Observatory of Japan in Hilo, Hawaii.

These and other planned surveys will set the stage for weak lensing's coup de grâce next decade. Tyson leads a large U.S. team that is working on the Large Synoptic Survey Telescope (LSST), a project that has won top billing for ground-based astronomy from national review panels. A radical optical design of one 8.4-meter mirror and two other mirrors larger than 4 meters will open up a giant swath of sky—at least 7 square degrees—for LSST to see at once. Among many projects, LSST will discover 300,000 mass clusters and tighten the errors on cosmic parameters—such as the dark energy "equation of state," a measure of its physical cause—to about 2%, Tyson predicts. He hopes observations will begin by 2011.



Wide eye. The Large Synoptic Survey Telescope will look for dark-matter warping.

At about the same time, supernova researchers led by astrophysicist Saul Perlmutter of Lawrence Berkeley National Laboratory in California hope to launch the SuperNova Acceleration Probe (SNAP). The satellite, an ambitious proposal to study dark energy by tracing the expansion history of the universe more than 10 billion years into the past, will carry a wide-field 2-meter telescope ideal for measuring weak lensing as well. Current plans call for SNAP to devote 32 months to supernova searches and 5 months to a weak-lensing survey spanning at least 300 square degrees, Perlmutter says.

Lensing aficionados hope to avoid a battle for funding between the two expensive approaches. Research on the cosmic microwave background radiation showed that cleverly designed telescopes on the ground and on balloons could answer key questions. Then, the Wilkinson Microwave Anisotropy Probe satellite nailed the answers beyond doubt from the quiet of space. In a similar vein, outside observers think that both future lensing projects should proceed. Still, some believe that SNAP may yield the most stunning results. "We need to measure the shapes of galaxies as accurately as possible, and we have problems [doing that] from the ground," says Van Waerbeke of IAP. "But from space, it's just perfect."

That debate may sharpen as weak lensing becomes more widely known, but so will the basic shift in how we study the cosmos. "The universe is not those pinpoints of light we can see in the night," Tyson says. "It is in fact this dark side. In some sense, we are using what most people *thought* was the universe, namely radiation and light, as a tool to measure the real universe for the first time." As that door opens, we will grow accustomed to a warped universe where no shining object is quite as it appears.

–ROBERT IRION

NEWS

Dark Energy Tiptoes Toward the Spotlight

Discovered less than a decade ago, a mysterious antigravity force suffuses the universe. Physicists are now trying to figure out the properties of this "dark energy"—the blackest mystery in the shadiest realms of cosmology

It's the biggest question in physics: What is the invisible stuff blowing the universe apart? A decade ago, the idea of "dark energy" was a historical footnote, something Einstein concocted to balance his equations and later regretted. Now, thanks to observations of distant supernovae and the faint afterglow of the big bang, dark energy is weighing ever more heavily upon the minds of cosmologists. They now know that this mysterious "antigravity" force exists, yet nobody has a good explanation for what it might be or how it works.

That vexing state of affairs may be starting to change. Scientists are finally beginning to get the first tentative measurements of the properties of this ineffable force. It's a crucial endeavor, because the nature of dark energy holds the secret to the fate of the universe and might even cause its violent and sudden demise.

"We're off to a very good start," says Adam Riess, an astronomer at the University of California (UC), Berkeley, who hints that within the next few months, supernova observations will finally help scientists begin to shine light on dark energy.

The modern story of dark energy began in 1997 when supernova hunters such as Riess and Saul Perlmutter of Lawrence Berkeley National Laboratory in California shocked the scientific community by showing that the universe is expanding ever faster rather than slowing down as physicists expected. They based that conclusion on observations of large numbers of supernovae known as type Ia. Because every type Ia explodes in roughly the same way with roughly the same brightness, the astronomers could use characteristics of their light to determine how far away the supernovae are (which is equivalent to determining how old they are) and how fast they're moving. When they calculated how fast the universe had been expanding at various times in the past, the results were "a big surprise," says Perlmutter: The universe has been expanding faster and faster rather than slowing down.

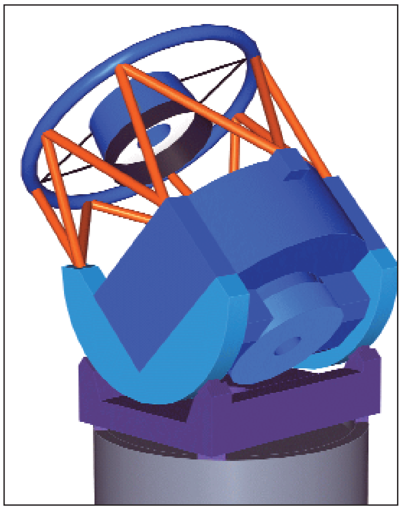
On the face of it, this was an absurd conclusion. As far as most physicists were concerned, only two big forces had shaped the universe. First, the energy of the big bang caused the early universe to expand very rapidly; then as the energy and matter in the universe condensed into particles, stars, and galaxies, the mutual gravitation of the mass started putting on the brakes.

The supernova data showed that something else has been going on. It is as if some mysterious antigravity force is making the fabric of the universe inflate faster than gravity can make it collapse (*Science*, 30 January 1998, p. 651). Observations of the cosmic microwave background radiation bolstered the case. By looking at the patchiness in the microwave radiation from the early universe, cosmologists could see that the universe as a whole is "flat": The fabric of spacetime has no curvature (*Science*, 28 April 2000, p. 595). Yet there is far too little matter in the universe to pull it into such a shape. There has to be an unknown energy—dark energy—suffusing the uni-

CREDIT: LSST CORP.

dark entities. "We would like to publish the first mass-selected object catalog [of dark-matter lenses] in a timely manner," says team leader Satoshi Miyazaki of the National Astronomical Observatory of Japan in Hilo, Hawaii.

These and other planned surveys will set the stage for weak lensing's coup de grâce next decade. Tyson leads a large U.S. team that is working on the Large Synoptic Survey Telescope (LSST), a project that has won top billing for ground-based astronomy from national review panels. A radical optical design of one 8.4-meter mirror and two other mirrors larger than 4 meters will open up a giant swath of sky—at least 7 square degrees—for LSST to see at once. Among many projects, LSST will discover 300,000 mass clusters and tighten the errors on cosmic parameters—such as the dark energy "equation of state," a measure of its physical cause—to about 2%, Tyson predicts. He hopes observations will begin by 2011.



Wide eye. The Large Synoptic Survey Telescope will look for dark-matter warping.

At about the same time, supernova researchers led by astrophysicist Saul Perlmutter of Lawrence Berkeley National Laboratory in California hope to launch the SuperNova Acceleration Probe (SNAP). The satellite, an ambitious proposal to study dark energy by tracing the expansion history of the universe more than 10 billion years into the past, will carry a wide-field 2-meter telescope ideal for measuring weak lensing as well. Current plans call for SNAP to devote 32 months to supernova searches and 5 months to a weak-lensing survey spanning at least 300 square degrees, Perlmutter says.

Lensing aficionados hope to avoid a battle for funding between the two expensive approaches. Research on the cosmic microwave background radiation showed that cleverly designed telescopes on the ground and on balloons could answer key questions. Then, the Wilkinson Microwave Anisotropy Probe satellite nailed the answers beyond doubt from the quiet of space. In a similar vein, outside observers think that both future lensing projects should proceed. Still, some believe that SNAP may yield the most stunning results. "We need to measure the shapes of galaxies as accurately as possible, and we have problems [doing that] from the ground," says Van Waerbeke of IAP. "But from space, it's just perfect."

That debate may sharpen as weak lensing becomes more widely known, but so will the basic shift in how we study the cosmos. "The universe is not those pinpoints of light we can see in the night," Tyson says. "It is in fact this dark side. In some sense, we are using what most people *thought* was the universe, namely radiation and light, as a tool to measure the real universe for the first time." As that door opens, we will grow accustomed to a warped universe where no shining object is quite as it appears.

–ROBERT IRION

NEWS

Dark Energy Tiptoes Toward the Spotlight

Discovered less than a decade ago, a mysterious antigravity force suffuses the universe. Physicists are now trying to figure out the properties of this "dark energy"—the blackest mystery in the shadiest realms of cosmology

It's the biggest question in physics: What is the invisible stuff blowing the universe apart? A decade ago, the idea of "dark energy" was a historical footnote, something Einstein concocted to balance his equations and later regretted. Now, thanks to observations of distant supernovae and the faint afterglow of the big bang, dark energy is weighing ever more heavily upon the minds of cosmologists. They now know that this mysterious "antigravity" force exists, yet nobody has a good explanation for what it might be or how it works.

That vexing state of affairs may be starting to change. Scientists are finally beginning to get the first tentative measurements of the properties of this ineffable force. It's a crucial endeavor, because the nature of dark energy holds the secret to the fate of the universe and might even cause its violent and sudden demise.

"We're off to a very good start," says Adam Riess, an astronomer at the University of California (UC), Berkeley, who hints that within the next few months, supernova observations will finally help scientists begin to shine light on dark energy.

The modern story of dark energy began in 1997 when supernova hunters such as Riess and Saul Perlmutter of Lawrence Berkeley National Laboratory in California shocked the scientific community by showing that the universe is expanding ever faster rather than slowing down as physicists expected. They based that conclusion on observations of large numbers of supernovae known as type Ia. Because every type Ia explodes in roughly the same way with roughly the same brightness, the astronomers could use characteristics of their light to determine how far away the supernovae are (which is equivalent to determining how old they are) and how fast they're moving. When they calculated how fast the universe had been expanding at various times in the past, the results were "a big surprise," says Perlmutter: The universe has been expanding faster and faster rather than slowing down.

On the face of it, this was an absurd conclusion. As far as most physicists were concerned, only two big forces had shaped the universe. First, the energy of the big bang caused the early universe to expand very rapidly; then as the energy and matter in the universe condensed into particles, stars, and galaxies, the mutual gravitation of the mass started putting on the brakes.

The supernova data showed that something else has been going on. It is as if some mysterious antigravity force is making the fabric of the universe inflate faster than gravity can make it collapse (*Science*, 30 January 1998, p. 651). Observations of the cosmic microwave background radiation bolstered the case. By looking at the patchiness in the microwave radiation from the early universe, cosmologists could see that the universe as a whole is "flat": The fabric of spacetime has no curvature (*Science*, 28 April 2000, p. 595). Yet there is far too little matter in the universe to pull it into such a shape. There has to be an unknown energy—dark energy—suffusing the uni-

CREDIT: LSST CORP.

verse. "The fact that these two teams came up with essentially the same result is why they are taken so seriously," says Alexei Filippenko of UC Berkeley. "With each year, it's taken more and more seriously."

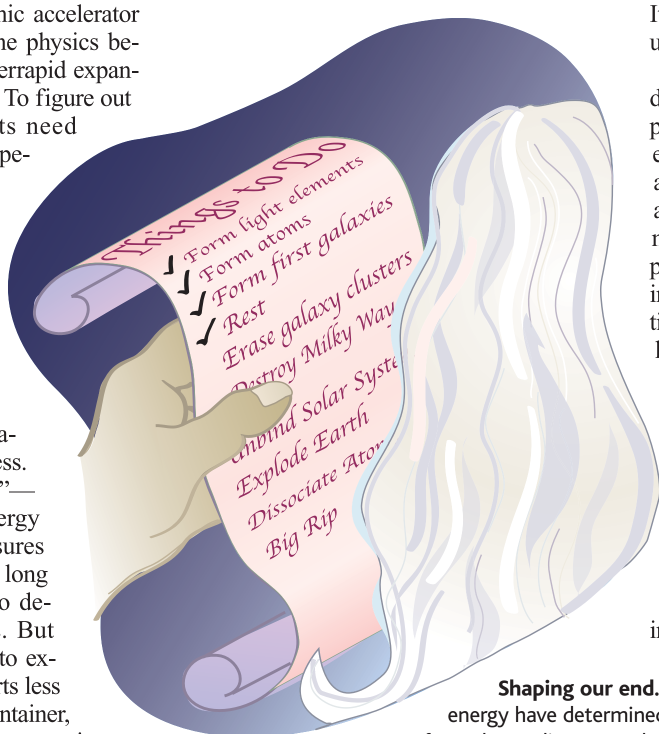
So, what is dark energy? Some theorists think it might be the energy latent in the vacuum itself. According to the rules of quantum mechanics, even empty space is seething with particles—particles that can exert pressure (*Science*, 10 January 1997, p. 158). It may be that the vacuum energy somehow is causing the fabric of spacetime to expand ever faster. Other physicists suspect that the foot on the cosmic accelerator might be a weaker form of the physics behind inflation, a period of superrapid expansion shortly after the big bang. To figure out what is going on, physicists need more information about the specific properties of dark energy.

Luckily, cosmologists and astronomers are finally beginning to get data that allow them to delve into those properties. One of the key targets is "*w*": the so-called equation of state of dark energy. "*w* is a parameter which will characterize the nature of dark energy," says Riess. "It tells you how squishy it is"—more precisely, how dark energy behaves under different pressures and densities. Physicists have long invoked similar parameters to describe the behavior of gases. But whereas a gas, when allowed to expand into a larger volume, exerts less pressure on the walls of its container, dark energy exerts *more* pressure as it expands. This counterintuitive property makes the value of *w* a negative number rather than a positive number.

In cosmological models, the "container" is the universe itself. At any given moment, its volume determines the pressure that drives the universe to expand. In theory, the pressure could have been affected by the volume in any of infinitely many ways, each writing the history of the universe in a slightly different manner. To find out which scenario we live in, physicists need to nail down how forcefully the dark energy is bearing down on the universe and whether the push has varied over time.

The key to that determination is *w*. If dark energy's pressure has been constant throughout the history of the universe, *w* is –1. If the properties of dark energy have been changing over time, as various "quintessence" theories suggest, *w* lies between 0 and –1 and might even change as time passes. According to Riess, unpublished supernova measurements by the Hubble Space Telescope and other sources indicate that *w* is about –1. "We should get the first very crude estimates of whether *w* is changing later this year," he adds.

However, the supernova results leave open a bizarre possibility. Earlier this year, physicist Robert Caldwell of Dartmouth College in Hanover, New Hampshire, and his colleagues investigated what happens if *w* is less than –1, for example, if it's –1.1 or –1.2 or –2. Physicists had shied away from such values, because they make theoretical equations start spewing out ugly infinities and other logical inconsistencies. But Caldwell's group didn't flinch. "Interesting things happen as dark energy becomes more and more repulsive," says Caldwell.



Shaping our end. The properties of dark energy have determined the universe's history so far and may dictate an alarming denouement.

ILLUSTRATION: A. STONEBRAKER/*SCIENCE*

"Interesting" is putting it mildly: The universe dies a horrible death. The ever-strengthening dark energy makes the fabric of the universe expand ever faster and things fall apart. In a few billion years, galaxy clusters disintegrate. The galaxies' mutual pull is overwhelmed by the dark energy, and they spin away from each other in ever-widening gyres. Several hundred million years later, galaxies themselves, including our own Milky Way, fling themselves to pieces. Solar systems and planets spin into fragments. Even atoms lose control of their electrons, and then atomic nuclei get torn apart and protons and neutrons shatter under the enormous expanding pressure. "Space becomes unstable," says Riess. The universe ends in a "big rip," a cataclysm where all matter gets shredded by the ever-stretching fabric of spacetime.

Although few physicists favor the big-rip scenario, nobody can rule it out a priori. In fact, some big-rip values of *w* could explain the supernova data pretty well. "Apart from distaste, there's no other reason and no observations pushing you to a *w* greater than –1," says Caldwell. Riess agrees: "Some of the values look like a good fit, –1.1 or –1.2." Unfortunately, although supernova data are rapidly narrowing down the possible values of *w* greater than –1, they don't shed nearly as much light on the regime below *w* of –1. It will be a while before physicists can figure out whether the big rip awaits us.

In the meantime, other scientists are using distant supernovae to figure out another aspect of dark energy's history. Because dark energy gets relatively stronger as it expands and the force of gravity gets relatively weaker as matter gets more diffuse, they reason, there must have been a time when dark energy's expansionist push was weaker than the contracting force of gravity. Cosmologists think the tipping point occurred when the universe was less than about 4 billion years old. Before then, the expansion of the universe must have been slowing—just as physicists used to assume it was doing today.

By pinpointing when the era of slowing gave way to the era of speeding up, Riess says, supernova hunters can test whether dark energy really behaves as theorists assume it does—or whether it defies all expectations. "And what's exciting is that we have data in the can now" that might pinpoint that time, says Riess. According to Caldwell, figuring out when the deceleration switched to acceleration might yield even more information about the nature of dark energy than *w* can: It will be a relatively sensitive probe to the strength of the energy. And although Riess and Perlmutter haven't released their full data sets yet, Filippenko says that there is a "hint" in the data of this ancient deceleration before the acceleration.

These are baby steps into a new realm of physics that was entirely obscure until a few years ago—and scientists are just beginning to figure out its properties. "I'd love to be able to take a lump of dark energy and see what happens when you knock it about, squish it, drop it on the floor," says Campbell. But short of that, observations of supernovae and eventually the evolution of distant galaxy clusters and galaxies will begin to pull back the veil over dark energy. Until then, dark energy will likely be the darkest mystery in a very dark universe.

–CHARLES SEIFE

The Dark Age of the Universe

Jordi Miralda-Escudé[1,2,3]

The Dark Age is the period between the time when the cosmic microwave background was emitted and the time when the evolution of structure in the universe led to the gravitational collapse of objects, in which the first stars were formed. The period of reionization started with the ionizing light from the first stars, and it ended when all the atoms in the intergalactic medium had been reionized. The most distant sources of light known at present are galaxies and quasars at redshift $z \cong 6$, and their spectra indicate that the end of reionization was occurring just at that time. The Cold Dark Matter theory for structure formation predicts that the first sources formed much earlier.

It was only about 75 years ago when Edwin Hubble discovered that we live in a universe of galaxies in expansion. At about the same time, Alexander Friedmann used the cosmological principle (the assumption that the universe can be approximated on large scales as homogeneous and isotropic) to write down the basic equations governing the structure and evolution of the universe in the Big Bang model, starting from Einstein's theory of General Relativity. By the end of the 20th century, much evidence had accumulated showing that the early universe was close to homogeneous, even on the small scales of the present galaxies. The fundamental question is how the universe went from this initial nearly homogeneous state to the present-day extremely complex form, in which matter has collapsed into galaxies and smaller structures.

I will review the history of the universe from the time of emission of the cosmic microwave background (CMB) to the time when the first objects collapsed gravitationally. An overview of these events will be described, with respect to the time and the redshift at which they take place (Fig. 1). Cosmologists generally use the redshift z to designate a cosmic epoch. The quantity $1 + z$ is the factor by which the universe has expanded from that epoch to the present time and is also the factor by which the wavelength of the light emitted by any object at that epoch and reaching us at the present time has been stretched, owing to the expansion of the universe.

The Cold Dark Matter Model

Cosmological observations can be accounted for by the Cold Dark Matter (CDM) model [see (*1–3*) for reviews]. The model assumes that in addition to ordinary matter made of protons, neutrons, and electrons (usually referred to as baryonic matter in cosmology), there is also dark matter, which behaves as a collection of collisionless particles having no interactions other than gravity and which was initially cold (that is, the particles had a very small velocity dispersion). Observations have confirmed the existence of dark matter in galaxy halos and clusters of galaxies [e.g., (*4–9*)]. The intensity fluctuations of the CMB (the relic radiation that is left over from the epoch when the universe was nearly homogeneous and in thermal equilibrium) also reveal that for these primordial, small-amplitude fluctuations to have grown into the present galaxies, clusters, and large-scale structures of the universe through gravitational evolution, the presence of dark matter is required. More recently, another component has been identified, called dark energy, which has become the dominant component of the universe at the present epoch and is causing an acceleration of the expansion of the universe (*10*, *11*).The Wilkinson Microwave Anisotropy Probe (WMAP) (*12*, *13*) showed that the baryonic matter accounts for only ~17% of all matter, with the rest being the dark matter, and has confirmed the presence of the dark energy (*14*, *15*). Although the CDM model with the added dark energy agrees with many observations, cosmologists have no idea what the nature of the dark matter and the dark energy may be, and why this matter and energy should have comparable densities at the present time.

Nevertheless, as the parameters of this CDM model are measured more precisely, the predictions for the number of objects of different mass that should be gravitationally collapsing at every epoch in the universe have become more robust. Bound objects form when the primordial fluctuations reach an amplitude near unity, entering the nonlin-

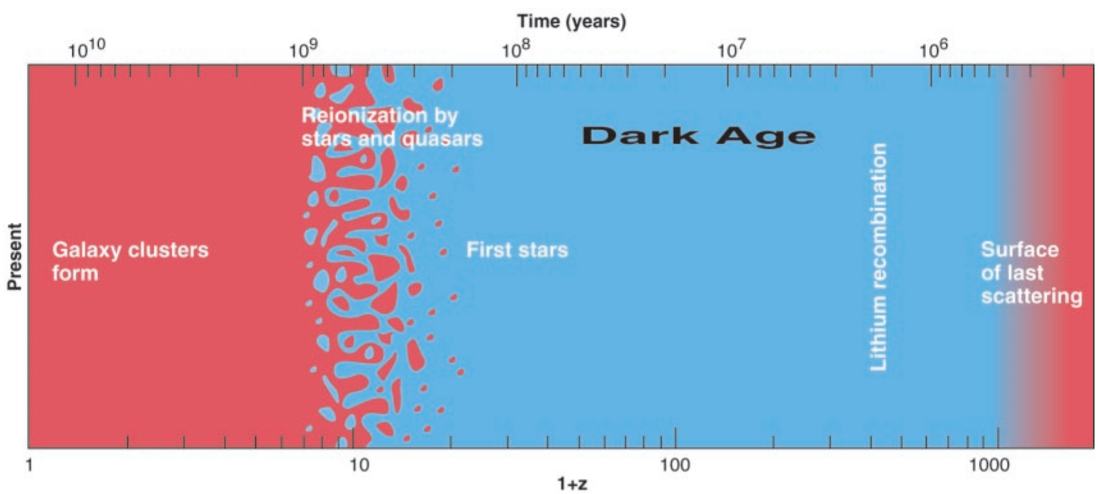


Fig. 1. Overview of the main events discussed in this review, with the top axis showing the age of the universe and the bottom axis the corresponding redshift, for the currently favored model (same parameters as in Fig. 2). Blue represents atomic regions, and red, ionized regions. Matter in the universe recombined in a homogeneous manner at $z \cong 1200$. Later, when the first stars formed and emitted ionizing radiation, ionized regions formed around the sources that eventually overlapped, filling all of space. The size of the HII regions should be much smaller on the redshift scale than shown here and is drawn only for illustration.

[1]Department of Astronomy, The Ohio State University, Columbus, OH 43210, USA. [2]Institute for Advanced Study, Princeton, NJ 08540, USA. [3]Institut d'Estudis Espacials de Catalunya/ICREA, Barcelona, Spain. E-mail: jordi@astronomy.ohio-state.edu

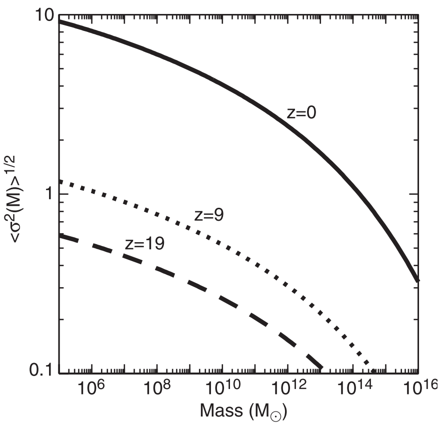


Fig. 2. The solid line shows the present time ($z = 0$), linearly extrapolated rms fluctuation $< (\delta M/M)^2 >^{1/2}$ of the mass enclosed in a region that contains an average mass M, expressed in the horizontal axis in units of solar masses. The other two curves are for $z = 9$ and $z = 19$, when the universe was about 500 million and 200 million years old, respectively. Fluctuations grow with time, and when they reach an amplitude near unity at some scale, nonlinear formation of halos takes place, and small halos merge into larger ones as progressively larger scales undergo collapse. The flat CDM model with cosmological constant assumed here has the following parameters: Hubble constant $H_0 = 70$ km s^{-1} Mpc^{-1}, matter density $\Omega_{m0} = 0.3$, baryon density $\Omega_b = 0.043$, amplitude of fluctuations $\sigma_8 = 0.9$, and primordial spectral index $n = 0.93$.

ear regime. The power spectrum of the fluctuations can be represented in terms of the root-mean-square (rms) fluctuation of the mass, δM, enclosed by a sphere of radius R, which on average has a mass M, equal to its volume times the mean density of the universe. The linearly extrapolated rms fluctuation $\delta M/M$ is shown in Fig. 2 as a function of M for the CDM model, at the present time ($z = 0$) and at redshifts $1 + z = 10$ and $1 + z = 20$. Note that linear fluctuations grow gravitationally in proportion to $(1 + z)^{-1}$, except at $z \lesssim 1$, when the dark energy starts to dominate (*16*). At the present time, fluctuations are typically of order unity on scales containing masses $\sim 10^{14}$ $M_\odot$ (where $M_\odot$ is solar mass), corresponding to galaxy groups. At the epoch $z = 9$, typical fluctuations were collapsing on much smaller scales of $M \sim 10^6$ $M_\odot$. Because the probability distribution of the mass fluctuation on any given region is Gaussian, there should be rare regions in the universe with a density fluctuation of several times the variance that will correspondingly be able to collapse earlier. For example, our Milky Way galaxy may have formed from the collapse of a 10^{12} $M_\odot$ halo from a 1σ fluctuation at $z \cong 1$, but at $z = 5$ halos of the same mass were already forming from 3σ fluctuations. On a scale of $10^6 M_\odot$, a 1σ fluctuation collapses at $z \cong 6$, and a 3σ fluctuation collapses at $z \cong 20$ (Fig. 3). Each object that forms has a velocity dispersion v determined by its mass and the size of the region from which it collapsed, $v^2 \sim GM/R$, and a corresponding virialized temperature of the gas, $kT_{vir} = (\mu m_H)\, v^2$, where μ is the mean particle mass in units of the hydrogen mass m_H. It is this virialized temperature that determines the physics of the rate at which gas can cool to form stars. This prediction of the number of objects that were forming at each z forms the basis for our ideas on the end of the Dark Age, the formation of the first stars, and the reionization.

The Dark Age

At very high z, the universe was practically homogeneous, and the temperature of matter and radiation dropped as the universe expanded. Atoms formed at $z \cong 1100$ when the temperature was $T = 3000$ K, a low enough value for the plasma to recombine. At this epoch of recombination, the CMB filled the universe with a red, uniformly bright glow of blackbody radiation, but later the temperature dropped and the CMB shifted to the infrared. To human eyes, the universe would then have appeared as a completely dark place. A long period of time had to pass until the first objects collapsed, forming the first stars that shone in the universe with the first light ever emitted that was not part of the CMB (Fig. 1). The period of time between the last scattering of the CMB radiation by the homogeneous plasma and the formation of the first star has come to be known as the Dark Age of the universe (*17*).

Observations provide detailed information on the state of the universe when the CMB radiation was last scattered at $z \cong 1100$, and we have also observed galaxies and quasars up to $z \cong 6.5$ (*18–21*). The theory suggests that the first stars and galaxies should have formed substantially earlier, so we can expect to discover galaxies at progressively higher z as technology advances and fainter objects are detected. However, beyond a z of 10 to 20, the CDM theory with Gaussian fluctuations predicts that the dark matter halos that can host luminous objects become extremely rare, even for low-mass halos (Fig. 2). Discovering any objects at $z \gtrsim 20$ should become exceedingly difficult as we reach the period of the Dark Age. During the Dark Age, before the collapse of any objects, not much was happening at all. The atomic gas was still close to homogeneous, and only a tiny fraction of it formed the first molecules of H_2, HD, and LiH as the temperature cooled down [e.g., (*22*, *23*)]. One of the few suggested ideas for an observational probe of the Dark ge is to detect secondary anisotropies on the CMB that were imprinted by Li atoms as they recombined at $z \cong 400$ through the resonance line at 670.8 nm, which would be redshifted to the far-infrared today, making it difficult to observe because of the foreground emission by dust (*24*, *25*).

How Did the First Stars Form?

The Dark Age ended when the first stars were formed. In order to form stars, the atomic gas must be able to follow the collapse of dark matter halos. This happens when the halo mass is above the Jeans mass of the gas (*26*) at the virialized temperature and density of the intergalactic medium, a condition that is fulfilled when $T_{vir} \gtrsim 100$ K (*1*, *27*). In halos with lower temperature, the gas pressure is sufficient to prevent the gas from collapsing. In addition, there must be a radiative cooling mechanism for the gas to lose its energy and concentrate to ever-higher densities in the halo centers until stellar densities are reached; without cooling, the gas reaches hydrostatic equilibrium in the halo after the gravitational collapse and stays at a fixed density without forming stars. The ability of the gas to cool depends on T_{vir} and the chemical composition of the gas. T_{vir} was low for the first objects that formed and then it increased rapidly with time (Fig. 3). The primordial gas in the first halos was mainly composed of atomic H and He. Atomic H induces radiative cooling only when $T_{vir} > 10^4$ K, when collisions can excite and ionize H atoms (*28*); the gas can then readily con-

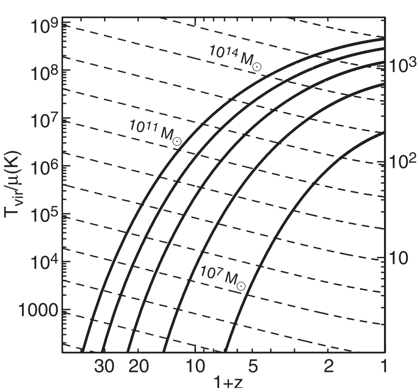


Fig. 3. The velocity dispersion v (right axis) or virialized temperature T_{vir} divided by the mean particle mass μ in units of the hydrogen mass (left axis; $\mu = 0.6$ for ionized matter and $\mu = 1.2$ for atomic matter) of halos collapsing from a 1σ fluctuation (of amplitude shown in Fig. 2) is shown as a function of redshift, as the lowest thick solid line. At every redshift, the fluctuation amplitude required for nonlinear collapse is reached at progressively larger scales, forming halos of increasing mass and velocity dispersion. The higher solid thick lines indicate halos collapsing from (2,3,4,5)-σ fluctuations, which form increasingly rare objects from a Gaussian distribution of fluctuations. The dashed lines indicate halos of constant mass, and are separated by a factor 10 in mass, with values indicated for three lines. Objects of fixed mass have increasing velocity dispersion as they form at higher redshift from a more rare, higher amplitude fluctuation because their size R is smaller.

tract to form galaxies. In the intermediate range 100 K $< T_{\rm vir} <$ 10^4 K, the gas settles into halos but atomic cooling is not available and, in the absence of the heavy elements that were formed only after massive stars ejected their synthesized nuclei into space, the only available coolant is H_2. Because two hydrogen atoms cannot form a molecule by colliding and emitting a photon, only a small fraction of the gas in these first objects could become H_2 via reactions involving the species H^- and $H_2{}^+$, formed by the residual free electrons and protons left over from the early universe (*29–31*), limiting the rate at which the gas could cool. Simulations (*32–37*) have shown that the first stars form in halos with $T_{\rm vir} \cong 2000$ K and mass $\sim 10^6\ M_\odot$; at lower temperatures, the rotational transitions of H_2 do not provide sufficient cooling for the gas to dissipate its energy. The slow cooling in these first objects leads to the formation of a central core with a mass of 100 to 1000 $M_\odot$ of gas cooled to ~200 K, and this core may form a massive star.

As soon as the first stars appeared, they changed the environment in which they were formed, affecting the formation of subsequent stars. Massive stars emit a large fraction of their light as photons that can ionize H (with energies greater than 13.6 eV), creating HII regions and heating the gas to $T \cong 10^4$ K. While these ionizing photons are all absorbed at the HII region boundaries, in the vicinity of the stars that emit them, photons with lower energy can travel greater distances through the atomic medium and reach other halos. Ultraviolet photons with energies above 11 eV can photodissociate H_2, and this can suppress the cooling rate and the ability to form stars in low-mass halos that are cooling by H_2 when they are illuminated by the first stars (*38*). The importance of this suppresion and other effects are being debated (*37*, *39–43*). Such effects might imply that the first massive stars formed through the radiative cooling of H_2 were a short-lived and self-destructive generation, because their own light might destroy the molecules that made their formation possible.

When some of these massive stars end their lives in supernovae, they eject heavy elements that pollute the universe with the ingredients necessary to form dust and planets (*44*). In a halo containing $10^6\ M_\odot$ of gas, the photoionization and supernova explosions from only a few massive stars can expel all the gas from the potential well of the halo (*45*). For example, the energy of 10 supernovae (about 10^{52} erg) is enough to accelerate $10^6\ M_\odot$ of gas to a speed of 30 km s^{-1}, which will push the gas out of any halo with a much lower velocity dispersion. The expelled gas can later fall back as a more massive object is formed by mergers of pre-existing dark matter halos. The next generation of stars can form by cooling provided by heavy elements (*46*), or by atomic H when $T_{\rm vir} > 10^4$ K. Abundances of heavy elements as low as 1000 times smaller than that of the sun can increase the cooling rate over that provided by H_2 and can also cool the gas to much lower temperatures than possible with H_2 alone, reducing the Jeans mass and allowing for the formation of low-mass stars (*47–49*).

A fascinating probe to these early events is provided by any stars that formed at that time with mass ~0.8 solar masses, which could be observed at the present time in our Galaxy's halo as they start ascending the red giant branch (*50*) if the halos in which they formed were later incorporated into the Milky Way by mergers. These stars should carry the signature of the elements synthesized by the first supernovae (*51*, *52*).

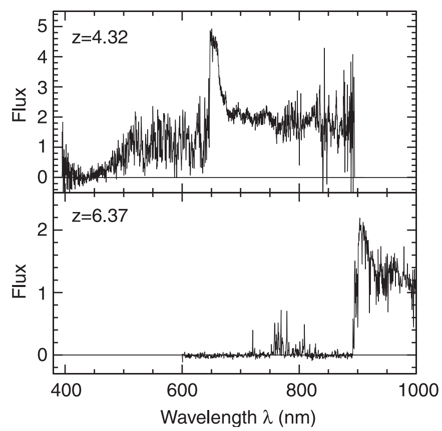


Fig. 4. Spectra of the Sloan Digital Sky Survey quasars J0019-0040 at $z = 4.32$, and J1148+5251 at $z = 6.37$. The flux is shown in units of 10^{-17} erg cm^{-2} s^{-1} as a function of wavelength. The peak of the spectra is the redshifted broad Lyα emission line of the quasars. Absorption by intervening hydrogen is seen at shorter wavelengths. At redshifts below 6 ($\lambda \lesssim 850$ nm), the medium is photoionized and the very small fraction of hydrogen that is atomic produces a partial, strongly fluctuating absorption reflecting the density variations of the intergalactic medium. At $z \cong 6$, the absorption suddenly becomes complete. This probably indicates the end of reionization. At $z >$ 6, the medium still contained atomic patches that are highly opaque to Lyα photons, and, even in the reionized regions, the ionizing background intensity was too low to reduce the neutral fraction to the very low values required for Lyα transmission. This figure is reproduced from [(*18*) fig. 3] and [(*19*) fig. 6].

When Did the First Star Form?

Because the primordial density fluctuations in the universe are random, the question of when the very first star formed does not have a simple answer. The time when the first halo with $T_{\rm vir} = 2000$ K collapsed depends on how rare a fluctuation we are willing to consider. A 5σ fluctuation in the density field can lead to the collapse of a halo and the formation of a star at $z \cong 30$ (Fig. 3). A more specific question we can ask is: From a random location in the universe, when would the first light from a star have been observed? Because an observer receives light only from the past light-cone (*53*), the further away one looks, the greater the volume that can be surveyed (and hence a more rare, higher-amplitude fluctuation can be found) but also the further back into the past one observes, which requires an even higher primordial density fluctuation to form a star. By requiring that just one collapsed halo with $T_{\rm vir} >$ 2000 K is observed on the past light-cone [and for the CDM model (Fig. 3)], a hypothetical observer located at a random place, after having experienced the dark age, would have seen the first star appear in the sky at $z \cong 38$ (*54*), when the universe was 75 million years old. This star would have formed from a 6.3σ fluctuation (with a probability of only $\sqrt{2/\pi}\int_{6.3}^{\infty} e^{-x^2/2}\,dx \cong 3\times10^{-10}$, implying that a volume containing a mass of $10^6\ M_\odot/3\times10^{-10} \cong 3\times10^{15}\ M_\odot$ would need to be searched to find one halo of $10^6 M_\odot$ at this early time). Soon after that first star, many more would have appeared forming from less rare fluctuations.

Because we can now see the very first stars that formed in the universe out to a very large distance on our past light-cone, we can survey a much larger volume than could the overjoyed observer at $z \cong 38$ at the sight of the first star. With this larger volume, the highest z star on the sky should be one formed from an 8σ fluctuation at $z \cong 48$ (*54*). Although this first star would be too faint to detect with current technology, brighter sources can pave the way to discover more primitive objects than the presently known most distant galaxies at $z \cong 6.5$. Perhaps we may discover more objects at higher z than expected in the CDM model, for example due to the presence of non-Gaussian primordial fluctuations on small scales [e.g., (*55*)].

The Reionization of the Universe

The most important effect that the formation of stars had on their environment is the reionization of the gas in the universe. Even though the baryonic matter combined into atoms at $z \cong 1100$, the intergalactic matter must have been reionized before the present. The evidence comes from observations of the spectra of quasars. Quasars are extremely luminous objects found in the nuclei of galaxies that are powered by the accretion of matter on massive black holes (*56*). Because of their high luminosity, they are used by cosmologists as lamp posts allowing accurate spectra to be obtained, in which the analysis of absorption lines provides information on the state of the intervening intergalactic matter. The spectra of quasars show the presence of light at wavelengths shorter than the Lyman-alpha (Lyα) emission line of H. If the intergalactic medium is atomic, then any photons emitted at wavelengths shorter than Lyα (121.6 nm) would be scattered by H at some point on their journey to us, when their wavelength is redshifted to the Lyα line. The mean density of H in the universe, when it is all in atomic form, is enough to provide a scattering optical depth as large as $\sim 10^5$ (*57*). The suppression of the flux at wavelengths shorter than the Lyα emission line is called the Gunn-Peterson trough.

In quasars at $z < 6$, the Gunn-Peterson trough is not observed. Instead, one sees the flux partially absorbed by what is known as the Lyα forest: a large number of absorption lines of different strength along the spectrum (Fig. 4). The H atoms in the intergalactic medium producing this absorption are a small fraction of all of the H, which is in photoionization equilibrium with a cosmic ionizing background produced by galaxies and quasars (*58*). The absorption lines correspond to variations in the density of the intergalactic matter. The observation that a measurable fraction of Lyα flux is transmitted through the universe implies that, after $z = 6$, the entire universe had been reionized.

However, recently discovered quasars (*19*, *59*, *60*) show a complete Gunn-Peterson trough starting at $z \cong 6$ (Fig. 4). Although the lack of transmission does not automatically imply that the intervening medium is atomic (because the optical depth of the atomic medium at mean density is $\sim 10^5$, and so even an atomic fraction as low as 10^{-3} produces an optical depth of ~ 100, which implies an undetectable transmission fraction), analysis of the Lyα spectra in quasars at $z < 6$ (*61*, *62*) indicates that the intensity of the cosmic ionizing background increased abruptly at $z \cong 6$. The reason for the increase has to do with the way in which reionization occurred. Ionizing photons in the far-ultraviolet have a short mean free path through atomic gas in the universe, so they are generally absorbed as soon as they reach any region in which the gas is mostly atomic. Initially, when the first stars and quasars were formed, the ionizing photons they emitted were absorbed in the high-density gas of the halos hosting the sources. The intergalactic medium started to be reionized when sufficiently powerful sources could ionize all the gas in their own halos, allowing ionizing photons to escape. The reionization then proceeded by the expansion of ionization fronts around the sources (Fig. 5), separating the universe into ionized bubbles and an atomic medium between the bubbles (*63*). The ionized bubbles grew and overlapped, until every low-density region of the universe was reionized; this moment defines the end of the reionization period. High-density regions that do not contain a luminous internal source can remain atomic because the gas in them recombines sufficiently fast, and they can self-shield against the external radiation. When the ionized bubbles overlap, photons are free to travel for distances much larger than the size of a bubble before being absorbed, and the increase in the mean free path implies a similar increase in the background intensity. The exact way in which the background intensity should increase at the end of reionization, depending on the luminosity function and spatial distribution of the sources, has not yet been predicted by theoretical models of reionization [e.g., (*64*)], but a rapid increase in the mean free path should, if present, tell us the time at which the reionization of the low-density intergalactic medium was completed.

The observational pursuit of the reionization epoch may be helped by the optical afterglows of gamma-ray bursts, which can shine for a few minutes with a flux that is larger than even the most luminous quasars (*65–70*), probably due to beaming of the radiation. Because gamma-ray bursts may be produced by the death of a massive star, they can occur even in the lowest-mass halos forming at the earliest times, with fixed luminosities. Among other things, the absorption spectra of gamma-ray burst optical afterglows might reveal the damped Lyα absorption profile of the H in the intervening atomic medium (*68*) and absorption lines produced by neutral oxygen (which can be present in the atomic medium only, before reionization) ejected by massive stars (*71*, *72*).

Electron Scattering of the CMB by the Reionized Universe

Reionization made most of the electrons in the universe free of their atomic binding, and able to scatter the CMB photons again. Before recombination at $z = 1100$, the universe was opaque, but because of the large factor by which the universe expanded from recombination to the reionization epoch, the electron Thompson scattering optical depth produced by the intergalactic medium after reionization, τ_e, is low. If the universe had reionized suddenly at $z = 6$, then $\tau_e \cong 0.03$. Because the fraction of matter that is ionized must increase gradually, from the time the first stars were formed to the end of reionization at $z = 6$, τ_e must include the contribution from the partially ionized medium at $z > 6$, and it must therefore be greater than 0.03. The sooner reionization started, the larger the value of τ_e.

The WMAP mission has measured τ_e from the power spectrum of the polarization and temperature fluctuations of the CMB. A model-independent measurement from the polarization-temperature correlation gives $\tau_e = 0.16 \pm 0.04$ (*73*), but a fit to the CDM model with six free parameters using both the correlation of temperature and polarization fluctuations found by WMAP, and other data gives $\tau_e = 0.17 \pm 0.06$ (*13*). An optical depth as large as $\tau_e = 0.16$ is surprising because it implies that a large fraction of the matter in the universe was reionized as early as $z \cong 17$, when halos with mass as low as $10^7\ M_\odot$ could collapse only from 3σ peaks, and were therefore still very rare (Fig. 3). The errors on τ_e will need to be reduced before we can assign a high degree of confidence to its high value (*74*).

What are the implications of a high τ_e if it is confirmed? Measurements of the emission rate at $z \cong 4$ from the Lyα forest show that to obtain $\tau_e > 0.1$, the emission rate would need to increase with z (*75*), and a large increase is required up to $z \cong 17$ to reach $\tau_e = 0.16$. In view of the smaller mass fraction in collapsed halos at this high z, it is clear that a large increase in the ionizing radiation emitted per unit mass is required from $z \approx 6$ to 17. Models have been proposed to account for an early reionization, based on a high emission

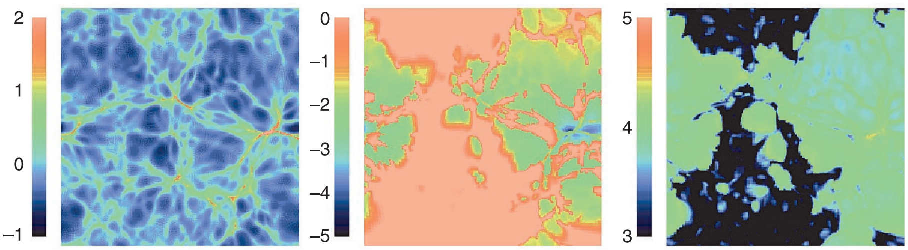


Fig. 5. Results of a simulation of the reionization of the intergalactic medium in a cubic box of co-moving side 4 h^{-1} Mpc, from (*64*) (Fig. 3B). The gas density (**left panel**), neutral fraction (**central panel**), and temperature (**right panel**) from a slice of the simulation are shown. The color coded values indicate the logarithms of the gas density divided by the mean baryon density, the neutral fraction, and the gas temperature in Kelvin, respectively. The simulation is shown at $z = 9$. The pink regions in the central panel are atomic, and the green regions are ionized. The sources of ionizing photons generally appear in halo centers where the gas density is high, but once the photons escape from the local high-density regions, the ionized bubbles expand most easily across the lowest density regions (compare left and central panels). The ionized regions are heated to about 10^4 K (see right panel), and they grow with time until they fill the entire universe at the end of reionization.

efficiency at high z (76–84). A possible reason for this high efficiency is that if the first stars that formed with no heavy elements were all massive (34, 36), they would have emitted as many as 10^5 ionizing photons per baryon in stars (85), many more than emitted by observed stellar populations (86–89). It is not clear, however, if enough of these massive stars can form in the first low-mass halos at $z > 17$, once the feedback effects of ultraviolet emission and supernovae (37, 38, 45) are taken into account. A different possibility might be that more objects than expected were forming at high z due to a fundamental change in the now favored CDM model.

The Future: the 21-cm Signature of the Atomic Medium

Many of the observational signatures of the epoch of reionization probe regions of the universe where stars have already formed and the medium has been reionized or polluted by heavy elements. But there is a way to study the undisturbed atomic medium. Nature turns out to be surprisingly resourceful in providing us with opportunities to scrutinize the most remote landscapes of the universe. The hyperfine structure of H atoms, the 21-cm transition due to the spin interaction of the electron and the proton, provides a mechanism to probe the atomic medium. When observing the CMB radiation, the intervening H can change the intensity at the redshifted 21-cm wavelength by a small amount, causing absorption if its spin temperature is lower than the CMB temperature, and emission if the spin temperature is higher. The spin temperature reflects the fraction of atoms in the ground and the excited hyperfine levels. The gas kinetic temperature cooled below the CMB temperature during the Dark Age owing to adiabatic expansion, although the spin temperature was kept close to the CMB temperature (90). When the first stars appeared in the universe, a mechanism for coupling the spin and kinetic temperature of the gas, and hence for lowering the spin temperature and making the H visible in absorption against the CMB, started to operate. The ultraviolet photons emitted by stars that penetrated the atomic medium were repeatedly scattered by H atoms after being redshifted to the Lyα resonance line, and these scatterings redistributed the occupation of the hyperfine structure levels (90–93), bringing the spin temperature down to the kinetic temperature and causing absorption. As the first generation of stars evolved, supernova remnants and x-ray binaries probably emitted x-rays that penetrated into the intergalactic medium and heated it by photoionization; gas at high density could also be shock-heated when collapsing into halos. The gas kinetic and spin temperatures could then be raised above the CMB, making the 21-cm signal observable in emission (93, 94). This 21-cm signal should reveal an intricate angular and frequency structure reflecting the density and spin temperature variations in the atomic medium (95–100). Several radio observatories will be attempting to detect the signal (101).

The observation of the 21-cm signal on the CMB will be a challenge, because of the long wavelength and the faintness of the signal. However the potential for the future is enormous: detailed information on the state of density fluctuations of the atomic medium at the epoch when the first stars were forming and the spin temperature variations that were induced by the ultraviolet and x-ray emission from the first sources are both encoded in the fine ripples of the CMB at its longest wavelengths.

References and Notes

1. G. R. Blumenthal, A. Dekel, J. R. Primack, M. J. Rees, *Nature* **311**, 517 (1984).
2. G. Efstathiou, in *Physics of the Early Universe*, J. A. Peacock, A. E. Heavens, A. T. Davies, Eds. (Hilger, Bristol, 1990) p. 361–463.
3. J. P. Ostriker, *Annu. Rev. Astron. Astrophys.* **31**, 689 (1993).
4. S. M. Faber, J. S. Gallagher, *Annu. Rev. Astron. Astrophys.* **2**, 135 (1979).
5. D. Zaritsky, S. D. M. White, *Astrophys. J.* **435**, 599 (1994).
6. T. A. McKay *et al.*, *Astrophys. J.* **571**, L85 (2002).
7. S. D. M. White, J. F. Navarro, A. E. Evrard, , C. S. Frenk, *Nature* **366**, 429 (1993).
8. S. Ettori, A. C. Fabian, *Mon. Not. R. Astron. Soc.* **305**, 834 (1999).
9. S. W. Allen, *Mon. Not. R. Astron. Soc.* **296**, 392 (1998).
10. S. Perlmutter *et al.*, *Astrophys. J.* **517**, 565 (1999).
11. A. G. Riess *et al.*, *Astrophys. J.* **560**, 49 (2001).
12. C. Bennett *et al.*, *Astrophys. J.* in press (e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302208).
13. D. N. Spergel *et al.*, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302209.
14. J. P. Ostriker, P. Steinhardt, *Science* **300**, 1909 (2003).
15. R. P. Kirshner, *Science* **300**, 1914 (2003).
16. Linear fluctuations grow in proportion to the scale factor, $(1 + z)^{-1}$, when matter dominates the energy density of the universe. The growth is being suppressed at present by the dark energy, and it was also suppressed in the early universe when radiation dominated over matter, at $z \gtrsim 3500$.
17. An excellent, more extensive review of the dark age can be found in R. Barkana, A. Loeb [*Phys. Rep.* **349**, 125 (2001)].
18. X. Fan *et al.*, *Astron. J.* **122**, 2833 (2001).
19. X. Fan *et al.*, *Astron. J.*, **25**, 1649 (2003).
20. E. Hu *et al.*, *Astrophys. J.* **568**, L75 (2002).
21. K. Kodaira *et al. Publ. Astron. Soc. Jpn.* **55**, L17 (2003).
22. S. Lepp, J. M. Shull, *Astrophys. J.* **280**, 465 (1984).
23. F. Palla, D. Galli, J. Silk, *Astrophys. J.* **451**, 44 (1995).
24. A. Loeb, *Astrophys. J.* **555**, L1 (2001).
25. P. C. Stancil, A. Loeb, M. Zaldarriaga, A. Dalgarno, S. Lepp, *Astrophys. J.* **580**, 29 (2002).
26. The Jeans mass of gas at a certain temperature and density is the minimum mass required for gravity to overcome the pressure gradient and force the gas to collapse.
27. H. M. P. Couchman, M. J. Rees, *Mon. Not. R. Astron. Soc.* **221**, 53 (1986).
28. Atomic helium needs even more energy to make a transition to the first excited state than atomic hydrogen, and therefore it can induce cooling only at higher temperatures than hydrogen.
29. W. C. Saslaw, D. Zipoy, *Nature* **216**, 976 (1967).
30. P. J. E. Peebles, R. H. Dicke, *Astrophys. J.* **154**, 891 (1968).
31. M. Tegmark*et al.*, *Astrophys. J.* **474**, 1 (1997).
32. T. Abel, P. Anninos, M. L. Norman, Y. Zhang, *Astrophys. J.* **508**, 518 (1998).
33. T. Abel, G. L. Bryan, M. L. Norman, *Astrophys. J.* **540**, 39 (2000).
34. T. Abel, G. L. Bryan, M. L. Norman, *Science* **295**, 93 (2002).
35. V. Bromm, P. S.Coppi, R. B. Larson, *Astrophys. J.* **527**, L5 (1999).
36. V. Bromm, P. S.Coppi, R. B. Larson, *Astrophys. J.* **564**, 23 (2002)
37. N. Yoshida, T. Abel, L. Hernquist, N. Sugiyama, *Astrophys. J.* in press (e-Print available at http://xxx.lanl.gov/abs/astro-ph/0301645).
38. Z. Haiman, M. J. Rees, A. Loeb, *Astrophys. J.* **476**, 458 (1997).
39. Z. Haiman, M. J. Rees, A. Loeb, *Astrophys. J.* **467**, 522 (1996).
40. M. Ricotti, N. Y. Gnedin, J. M. Shull, *Astrophys. J.* **560**, 580 (2001).
41. M. Ricotti, N. Y. Gnedin, J. M. Shull, *Astrophys. J.* **575**, 33 (2002).
42. M. Ricotti, N. Y. Gnedin, J. M. Shull, *Astrophys. J.* **575**, 49 (2002).
43. M. E. Machacek, G. L. Bryan, T. Abel, *Mon. Not. R. Astron. Soc.* **338**, 273 (2003).
44. Carbon and heavier elements were not synthesized in the Big Bang, and were all produced in stellar interiors. These elements therefore were not present in the primordial gas, and their abundance increased as supernovae and stellar winds delivered the products of the stellar nuclear furnaces to interstellar space.
45. V. Bromm, N.Yoshida, L. Hernquist, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0305333.
46. Species such as OI, FeII, and CII are the main coolants of gas at $T < 10^4$ K in the interstellar medium of galaxies at the present time.
47. V. Bromm, A. Ferrara, P. S. Coppi, R. B. Larson, *Mon. Not. R. Astron. Soc.* **328**, 969 (2001).
48. J. Mackey, V. Bromm, L. Hernquist, *Astrophys. J.* **586**, 1 (2003).
49. R. Schneider, A. Ferrara, P. Natarajan, K. Omukai, *Astrophys. J.* **571**, 30 (2002).
50. Stars of 0.8 solar masses have a lifetime about equal to the present age of the universe, so that one formed in the early universe should be running out of hydrogen in its core at present and becoming a luminous red giant. More massive stars live shorter and less massive ones live longer.
51. Y.-Z. Qian, G. J. Wasserburg, *Astrophys. J.* **567**, 515 (2002).
52. N. Christlieb *et al.*, *Nature* 419, 904 (2002).
53. The past light-cone is the set of all events in the universe from which a light message would be reaching us just at the present time. The further away we look, the further back into the past we are observing.
54. This simple calculation of the z at which the first object visible to an observer would collapse from a very rare fluctuation is based on the approximation that the collapse takes place when the linearly extrapolated overdensity reaches a fixed value, and is known as the Press-Schechter model, which was first proposed in (102) and is described in detail in its modern use in (103–105). The redshifts given here for the first object that would become visible may be changed slightly by corrections to this approximate model, and by changes in the CDM power spectrum normalization and primordial spectral index.
55. P. P. Avelino, A. R. Liddle, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0305357.
56. M. C. Begelman, *Science* **300**, 1898 (2003).
57. J. E. Gunn, B. A. Peterson, *Astrophys. J.* **142**, 1633 (1965).
58. M. Rauch, *Annu. Rev. Astron. Astrophys.* **36**, 267 (1998).
59. R. H. Becker, *et al.*, *Astron. J.* **122**, 2850 (2001).
60. R. L. White, R. H. Becker, X. Fan, M. A. Strauss, *Astron. J.*, in press (e-Print available at http://xxx.lanl.gov/abs/astro-ph/0303476).
61. P. McDonald, J. Miralda-Escudé, *Astrophys. J.* **549**, L11 (2001).
62. X. Fan *et al.*, *Astron. J.* **123**, 1247 (2002).
63. J. Arons, D. W. Wingert, *Astrophys. J.* **177**, 1 (1972).
64. N. Y. Gnedin, *Astrophys. J.* **535**, 530 (2000).
65. C. W. Akerlof *et al.*, *Nature* **398**, 400 (1999).

66. P. Mészáros, M. J. Rees, *Astrophys. J.* **476**, 231 (1997).
67. B. Zhang, S.Kobayashi, P. Mészáros, *Astrophys. J.*, in press (e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302525).
68. J. Miralda-Escudé, *Astrophys. J.* **501**, 15 (1998).
69. D. Q. Lamb, D. Reichart, *Astrophys. J.* **536**, 1 (2000).
70. V. Bromm, A. Loeb, *Astrophys. J.* **575**, 111 (2002).
71. S. P. Oh, *Mon. Not. R. Astron. Soc.* **336**, 1021 (2002).
72. S. R. Furlanetto, A. Loeb, *Astrophys. J.* **579**, 1 (2002).
73. A. Kogut *et al.*, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302213.
74. The errors are 1σ and include the best estimate of the WMAP team of systematic uncertainties associated with foreground Galactic emission; see table 2 of (*73*).
75. J. Miralda-Escudé, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0211071.
76. S. Whyithe, A. Loeb, *Astrophys. J. Lett.*, in press (e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302297).
77. Z. Haiman, G. Holder, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302403.
78. G. Holder, Z. Haiman, M. Kaplinghat, L. Knox, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302404.
79. B. Ciardi, A. Ferrara, S. D. M. White, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302451.
80. R. S. Somerville, M. Livio, e-Print available at http://xxx.lanl.gov/abs/ astro-ph/0303017.
81. R. S. Somerville, J. S. Bullock, M. Livio, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0303481.
82. A. Sokasian, T. Abel, L.Hernquist, V. Springel, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0303098.
83. R. Cen, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0303236.
84. W. A. Chiu, X. Fan, J. P. Ostriker, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0304234.
85. The maximum theoretical upper limit for the number of ionizing photons that can be emitted by a star for every baryon it contains is obtained by dividing the fusion energy per baryon released by fusion to helium, 7 MeV, by the average energy of an ionizing photon, ~20 eV, which gives ~3 × 10^5. Massive stars with no heavy elements can fuse almost all their hydrogen content over their lifetime and are hot enough to emit most of their radiation as ionizing photons (*86–88*).
86. J. Tumlinson, J. M. Shull, *Astrophys. J.* **528**, L65 (2000).
87. V. Bromm, R. P. Kudritzki, A. Loeb, *Astrophys. J.* **552**, 464 (2001).
88. D. Schaerer, *Astron. Astrophys.* **382**, 28 (2002).
89. A. Venkatesan, J. Tumlinson, J. M. Shull, *Astrophys. J.* **584**, 621 (2003).
90. D. Scott, M. J. Rees, *Mon. Not. R. Astron. Soc.* **247**, 510 (1990).
91. S. A. Wouthuysen, *Astron. J.* **57**, 31 (1952).
92. G. B. Field, *Astrophys. J.* **129**, 551 (1959).
93. P. Madau, A. Meiksin, M. J. Rees, *Astrophys. J.* **475**, 492 (1997).
94. X. Chen, J. Miralda-Escudé, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0303395.
95. P. Tozzi, P.Madau, A. Meiksin, M. J. Rees, *Astrophys. J.* **528**, 597 (2000).
96. I. T. Iliev, P. R. Shapiro, A. Ferrara, H. Martel, *Astrophys. J.* **572**, 123 (2002).
97. C. L. Carilli, N. Y. Gnedin, F. Owen, *Astrophys. J.* **577**, 22 (2002).
98. S. R. Furlanetto, A. Loeb, *Astrophys. J.* **588**, 18 (2003).
99. B. Ciardi, P. Madau, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0303249.
100. U.-L. Pen, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0305387.
101. At present the Giant Metrewave Radio Telescope (www.gmrt.ncra.tifr.res.in) in India is already searching for high-z 21-cm signals, and more sensitive observatories being designed now are the Square Kilometer Array (www.skatelescope.org) and the Low Frequency Array (www.lofar.org).
102. W. H. Press, P. Schechter, *Astrophys. J.* **187**, 425 (1974).
103. J. R. Bond, S. Cole, G. Efstathiou, N. Kaiser, *Astrophys. J.* **379**, 440 (1991).
104. R. J. Bower, *Mon. Not. R. Astron. Soc.* **248**, 332 (1991).
105. C. Lacey, S. Cole, *Mon. Not. R. Astron. Soc.* **262**, 627 (1993).

106. I thank X. Fan and N. Gnedin for their permission to reproduce and their help in providing figures from their papers and P. Sieber for suggesting a good way to start this article. I also thank T. Abel, A. Loeb, M. Rees, and my referees for their comments.

REVIEW

New Light on Dark Matter

Jeremiah P. Ostriker[1] and Paul Steinhardt[2]

Dark matter, proposed decades ago as a speculative component of the universe, is now known to be the vital ingredient in the cosmos: six times more abundant than ordinary matter, one-quarter of the total energy density, and the component that has controlled the growth of structure in the universe. Its nature remains a mystery, but assuming that it is composed of weakly interacting subatomic particles, is consistent with large-scale cosmic structure. However, recent analyses of structure on galactic and subgalactic scales have suggested discrepancies and stimulated numerous alternative proposals. We discuss how studies of the density, demography, history, and environment of smaller-scale structures may distinguish among these possibilities and shed new light on the nature of dark matter.

The dark side of the universe first became evident about 65 years ago when Fritz Zwicky (*1*) noticed that the speed of galaxies in large clusters is much too great to keep them gravitationally bound together unless they weigh over 100 times more than one would estimate on the basis of the number of stars in the cluster. Decades of investigation confirmed his analysis (*2–5*), and by the 1980s, the evidence for dark matter with an abundance of about 20% of the total energy density of the universe was accepted, although the nature of the dark matter remained a mystery.

After the introduction of inflationary theory (*6*), many cosmologists became convinced that the universe must be flat and that the total energy density must equal the value (termed the critical value) that distinguishes a positively curved, closed universe from a negatively curved, open universe. Cosmologists became attracted to the beguiling simplicity of a universe in which virtually all of the energy density consists of some form of matter, about 4% being ordinary matter and 96% dark matter. In fact, observational studies were never really compliant with this vision. Although there was a wide dispersion in total mass density estimates, there never developed any convincing evidence that there was sufficient matter to reach the critical value. The discrepancy between observation and the favored theoretical model became increasingly sharp.

Dark energy came to the rescue when it was realized that there was not sufficient matter to explain the structure and nature of the universe (*7*). The only thing dark energy has in common with dark matter is that both components neither emit nor absorb light. On a microscopic scale, they are composed of different constituents. Most important, dark matter, like ordinary matter, is gravitationally self-attractive and clusters with ordinary matter to form galaxies. Dark energy is gravitationally self-repulsive and remains nearly uniformly spread throughout the universe. Hence, a census of the energy contained in galaxies would miss most the dark energy. So, by positing the existence of a dark energy component, it became possible to account for the 70 to 80% discrepancy between the measured mass density and the critical energy density predicted by inflation (*8–11*). Then, two independent groups (*12*, *13*) found evidence of the accelerated expansion of the universe from observations of supernovae, and the model with a dominant dark energy component, as illustrated in Fig. 1, became the concordance model of cosmology. The existence of dark energy has recently been independently confirmed by observations by the Wilkinson Microwave Anisotrope Probe [WMAP (*14*)] and has become accepted as an essential ingredient of the standard model (*15*).

Dark energy has changed our view of the role of dark matter in the universe. According to

[1]Department of Astrophysical Sciences, [2]Department of Physics, Princeton University, Princeton, NJ 08544, USA.

Evidence for Black Holes

Mitchell C. Begelman

Black holes are common objects in the universe. Each galaxy contains large numbers—perhaps millions—of stellar-mass black holes, each the remnant of a massive star. In addition, nearly every galaxy contains a supermassive black hole at its center, with a mass ranging from millions to billions of solar masses. This review discusses the demographics of black holes, the ways in which they interact with their environment, factors that may regulate their formation and growth, and progress toward determining whether these objects really warp spacetime as predicted by the general theory of relativity.

Black holes are places where gravity is so strong that nothing that enters them—not even light—can escape. Within a finite region surrounding the center of a black hole, all light rays and physically realizable trajectories of particles are directed inward. Space and time are so distorted that there is literally no way out.

Classical black holes are described by vacuum solutions of the Einstein field equations of general relativity, which imply that they contain a singularity beyond which trajectories cannot continue. The nature of the singularity is not fully understood, and it is probable that existing physical theories break down close to the singularity. But from an astrophysicist's point of view this hardly matters, because the singularity is hidden from view. It lies beyond the event horizon, the surface that bounds the region of no escape. The size of the horizon surrounding a nonrotating, uncharged black hole is characterized by the Schwarzschild radius, $R_S = 2GM_\bullet/c^2$, where G is Newton's constant of gravity, $M_\bullet$ is the mass of the hole, and c is the speed of light. R_S is about 3 km for a black hole of $M_\bullet = 1\ M_\odot$ (solar mass). Curiously, although it is believed that conditions can become chaotic and violently unpredictable as matter traverses the region inside the horizon, the exteriors of black holes are believed to behave in predictable and relatively simple ways. Black hole horizons are among the most comprehensively understood phenomena predicted by theory [see (*1*) for a nontechnical introduction to the theoretical properties of black holes; see (*2*) for an undergraduate-level presentation of mathematical aspects of black hole physics].

One aspect of this simplicity is the fact that any black hole can be characterized by its mass, angular momentum, and electric (or certain other types of quantum) charge. Charge may not be important in an astrophysical setting, so black holes effectively can be described by two parameters. Far from the horizon, black holes exert a gravitational influence just like any spherical body of the same mass. If the Sun collapsed symmetrically and became a black hole, there would be no change in Earth's orbit. Black holes have been discovered primarily by measuring their gravitational effects on distant bodies. By applying the Newtonian laws of gravity, astronomers have established that dark masses exist at the centers of galaxies (*3*). If these masses are not black holes, they would have to be dense clusters of very faint objects. In at least two cases, the concentration of such a cluster would have to be so extreme as to render this interpretation highly implausible (*4*). In other objects the cluster interpretation cannot be ruled out with certainty, although it seems improbable. Likewise, certain x-ray–emitting binary stars have been shown to contain a compact object too massive to be a neutron star or any other pressure-supported body, leaving a black hole as the only plausible alternative (*5, 6*). In most instances, these gravitational interactions are measured on scales several orders of magnitude larger than the putative event horizon. The argument that these objects are black holes is therefore indirect and is based on the elimination of other possibilities.

In many cases, however, we can observe radiation emitted by gas located just outside the horizon, or jets of plasma flowing outward from the region around the horizon, at close to the speed of light (*7*). A major objective of black hole research is to use such observations to map the structure of spacetime near the horizon, thus testing whether these objects have exactly the properties predicted by general relativity. For example, we are well on our way toward being able to measure the tornado-like twisting of spacetime attributable to a black hole's spin (Fig. 1).

The energy liberated by matter close to the horizon is prodigious and can markedly affect a black hole's surroundings. Large black holes at the centers of galaxies have been implicated in the energy balance of gas thousands of light years away, and they may play a crucial role in the galaxy formation process (*8*). The collapse of massive stellar cores to form stellar-mass black holes could be responsible for triggering certain types of gamma-ray bursts, the most luminous phenomenon known (*9*). Thus, black holes are not only astonishing physical entities in their own right, as well as laboratories for the most extreme conditions encountered in the post–big bang universe; they are also key players in phenomena with which we have long been familiar. To understand stars, galaxies, and the gas that lies between them, we must understand how black holes form, where they form, and how they affect their environments.

Black Hole Demographics

Two populations of black holes have been established observationally. Stellar-mass black holes are presumably the collapsed remnants of massive stars. Except for two recent candidates based on gravitational microlensing surveys (*10, 11*), all of the several dozen stellar-mass black hole candidates have been found in x-ray binaries, close binary systems in which matter is transferred from a normal star to the black hole and emits x-rays before disappearing beneath the horizon. The rapid variability and spectral peak at x-ray wavelengths imply that the emitting region is extremely compact (smaller than a few hundred km). To distinguish black hole candidates from accreting neutron stars, which are also found in x-ray binaries, it is necessary to establish that the accreting compact object is too massive to be a neutron star [i.e., more than about 2 or 3 $M_\odot$, where the range stems from uncertainties in the stiffness of matter at nuclear densities (*6*)]. This is done by applying Newtonian mechanics to measured and estimated parameters of the binary orbit, including the orbital period, orbital speed of the normal star, radius and spectral type of the normal star, and orbital inclination with respect to the line of sight (*5*). Usually, one obtains a lower limit on the mass of the compact object, which in a number of cases comfortably exceeds the maximum mass of a neutron star.

There is little hope of detecting the much more numerous stellar-mass black holes that are presumably isolated or in noninteracting binary systems, because they do not capture enough gas from the interstellar medium to be observable. Given that stars more massive than 20 to 40 $M_\odot$ probably form black holes (*12*), the number of stellar-mass black holes in the Milky Way Galaxy could be as large as 10 million to 1 billion. The range is due to

JILA, University of Colorado, 440 UCB, Boulder, CO 80309, USA. E-mail: mitch@jila.colorado.edu

uncertainties in the initial mass function and star formation history as well as the complexities of stellar collapse calculations.

Supermassive black holes are found at the centers of galaxies. They were first proposed to explain the prodigious energy outputs of quasars and are now understood to be the primary source of energy in all types of active galactic nuclei (AGN) (*13*). But although only one in 100 galaxies is active at any time, we now know that most if not all galaxies have supermassive black holes in their nuclei (*3*). These black holes are not accreting at a high rate, and in fact they are underluminous relative to expectations based on the availability of accretable gas (*14*). They are detectable only through their gravitational effects on distant stars and gas.

The strongest dynamical evidence for a supermassive black hole comes from the center of the Milky Way. Observations in the infrared, radio, and x-ray bands—which can pierce the thick dust obscuring the Galactic Center in the optical—reveal a compact, nonstellar source of radiation, Sgr A*, surrounded by a cluster of stars. By measuring stellar proper motions and radial velocities, it has been possible to infer that the position of Sgr A* coincides with a dark mass of 3×10^6 to 4×10^6 $M_\odot$. Orbits of several stars have been mapped (*15–17*), probing the black hole's gravitational field to within 1000 R_S (60 times the radius of Earth's orbit) (Fig. 2). The compact radio emission and x-ray flares produced by Sgr A* (*18*) presumably come from gas accreting onto the black hole at a low rate (10^{-10} to 10^{-7} $M_\odot$ year^{-1}, depending on assumptions).

The second strongest case for a supermassive black hole is equally remarkable. Maser emission produced by water molecules in the nucleus of the galaxy NGC 4258 delineates a nearly perfectly Keplerian thin disk (*19*). Observers using very-long-baseline radio interferometry have mapped the radial velocities, proper motions, and accelerations of the masers in three regions of the disk, overdetermining its kinematics. The rotation curve fits a Keplerian model so well that the black hole mass, $M_\bullet = 39 \times 10^6$ $M_\odot$, is the most accurate known. Although the masers probe a region of radius ~40,000 R_S, it would be difficult to explain the rotation curve by anything other than a single compact mass at the center of the disk (*4*).

Several dozen other nearby galaxies have yielded dynamical evidence for supermassive black holes from measurements of stellar velocity dispersions and rotation curves of stars or gas in the nucleus (*3*). The measurements typically probe regions of radius $>10^5$ R_S. Therefore, the evidence is less compelling than the case for the Galactic Center or NGC 4258 black hole, but is nonetheless strong because of the amount of mass that must otherwise be hidden in the nucleus.

The masses of supermassive black holes are strongly correlated with the properties of their host galaxies, indicating a connection between black hole formation and galaxy formation. The nature of the connection is unclear. Black hole masses correlate roughly linearly with the mass of the galaxy's bulge, and they appear to be even more tightly correlated with the bulge velocity dispersion, σ, exhibiting a proportionality $M_\bullet \propto \sigma^4$ (*20–22*) (Fig. 3).

Inventories of quasar light (*23*) suggest that supermassive black holes grew mainly by accreting gas rather than by mergers of smaller black holes. If even a few percent of the liberated energy emerged in kinetic form, there would have been enough energy to unbind the gas in the protogalactic host. Thus, supermassive black holes might have limited their own growth—or even the final mass of the host galaxy—by depositing this energy in their surroundings. Such feedback could have given rise to the observed correlations (*24, 25*). Indeed, similar feedback effects from accreting black holes may play an important role in the energetics of nearby clusters of galaxies (*8*). Alternatively, dynamical processes during galaxy formation

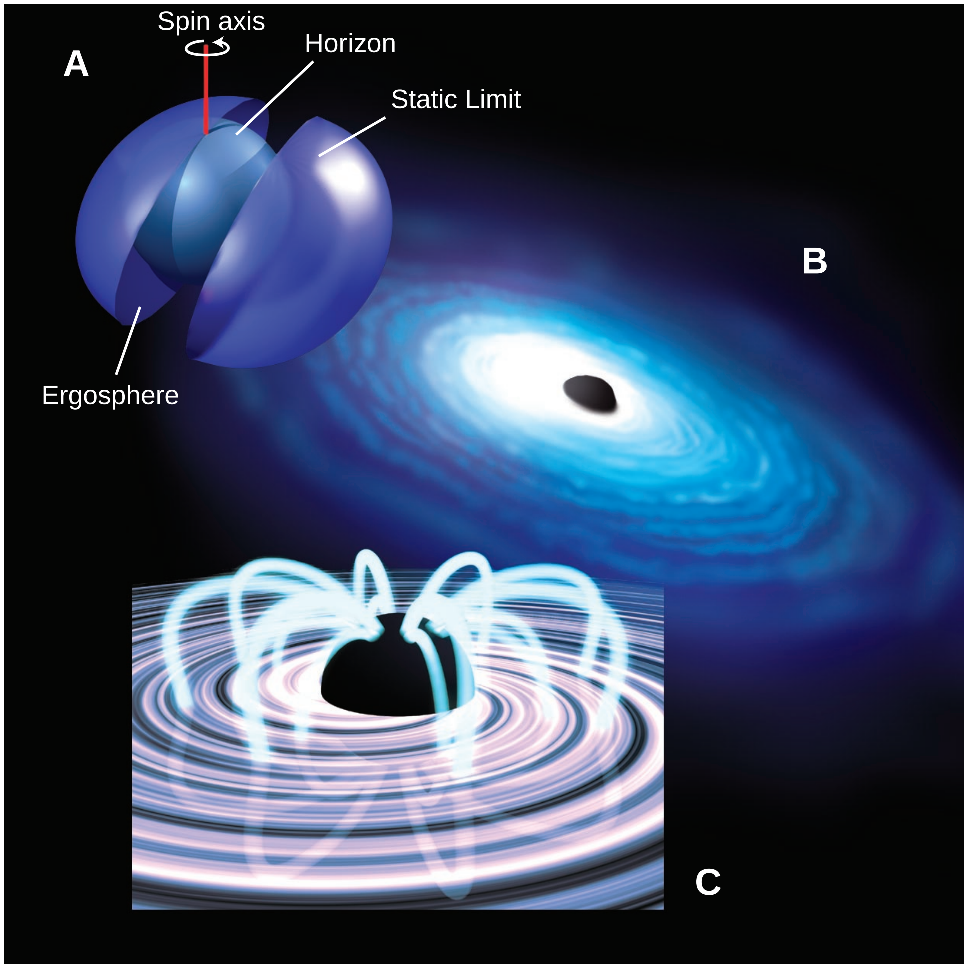


Fig. 1. Anatomy of a spinning black hole. (**A**) The event horizon, or surface of no-return, lies inside the static limit. Between the two surfaces, in the region known as the ergosphere, all trajectories must rotate in the same sense as the black hole. The ergosphere contains most of the black hole's spin energy, which can be extracted by magnetic fields. (**B**) Appearance of an accretion disk around a black hole. Gas on the left side is rotating toward the observer, and its emission is blueshifted; gas on the right side is rotating away from the observer, and its emission is redshifted. Emission from gas in front of the black hole is redshifted as a result of the gravitational redshift combined with the transverse Doppler shift. Actual size on the sky would be $<1 \times 10^{-6}$ arc sec (comparable to the size of newsprint seen from the distance of the Moon); such angular resolution could be attained with x-ray interferometers currently being designed. (**C**) Magnetic field is amplified inside the disk and erupts to form a corona. Magnetic field penetrating the ergosphere extracts energy from the black hole, which can accelerate a jet (if the field lines trail off into space) or enhance the emission from the disk (if the field lines connect to the disk, as shown here).

could have regulated the amount of matter that collected in the center and eventually formed the black hole (*26*).

There is intense interest in the possible existence of a third population of black holes, the so-called intermediate-mass black holes (IMBHs) (*27*). These would fill the gap in mass between stellar-mass and supermassive black holes. They could be the remnants of very massive (and hypothetical) Population III stars that formed from metal-free material in the early universe, or could have resulted from stellar mergers in dense star clusters. Their existence (or lack thereof) could tell us a lot about the conditions under which black holes formed and whether supermassive holes grew from much smaller ones, either by hierarchical mergers or by runaway accretion. As yet there is no solid evidence for such a population. It has been suggested that they could be associated with a class of x-ray binaries called ultraluminous x-ray sources (ULXs), which may require intermediate black hole masses in order to avoid disruption by radiation pressure (*28*). There is evidence that some ULXs may have cooler accretion disks than stellar-mass black holes, which supports the IMBH interpretation (*29*). But one cannot rule out alternative models that explain ULXs in terms of anisotropic emission (*30*) or radiation hydrodynamical effects (*31*) without recourse to masses higher than those of ordinary stellar-mass black holes. There are also candidates for IMBHs on the basis of velocity dispersions in globular clusters and gravitational microlensing surveys, but at present the evidence is not strong.

Interactions of Black Holes with Surrounding Matter

Black holes cannot swallow matter whose angular momentum per unit mass exceeds $\sim 2\ R_S c$. Astrophysically, this is a tiny amount of angular momentum. For example, in order to fall into a Sun-turned-black-hole, Earth would have to lose 99.99% of its orbital angular momentum. Contrary to popular belief, black holes are not cosmic vacuum cleaners. Gas orbiting under the gravitational influence of a black hole is thought to lose angular momentum to more distant gas via the magnetorotational instability (MRI) (*32*). In the presence of shear associated with orbital motion, a weak magnetic field is amplified to a fraction of the gas pressure within a few orbital periods. The resulting cross-correlation between the radial and tangential components of magnetic field causes a torque that transfers angular momentum outward. Thus, the magnetic stress behaves analogously to a shear viscosity, although with some very different (and not fully understood) detailed properties. As in any viscous fluid, the transport of angular momentum by MRI must be accompanied by dissipative heating and the outward transport of energy through the gas.

If the gas is able to radiate away the dissipated energy, it will settle into a geometrically thin Keplerian accretion disk (*33*). Gas in such a disk spirals inward gradually through a sequence of nearly circular orbits until it reaches the innermost stable circular orbit (ISCO). Once inside the ISCO, gas can fall into the black hole without any further loss of angular momentum. The ISCO is located well outside the event horizon, at 3 R_S, for a nonrotating hole and approaches the horizon (at $R_S/2$) for a rapidly spinning hole, provided that the gas is orbiting in the same sense as the hole's rotation. The total energy radiated by the disk is roughly half the gravitational potential energy at the ISCO; the other half is retained as kinetic energy of orbital motion, which disappears into the black hole along with the gas. Assuming that the stress is negligible inside the ISCO, this implies that each gram of accreting material radiates a fraction of its rest mass energy, ranging from 6% for a nonrotating hole to 42% for a hole near maximal rotation. The radiative efficiency could be even larger if magnetic stresses operate across the ISCO (*34*). These huge efficiencies, compared to the maximum ~1% efficiency of thermonuclear reactions, led to the early suggestions that only gravitational energy could power quasars. Indeed, thin accretion disks are the principal ingredients in models of x-ray binaries and luminous AGN such as quasars and Seyfert galaxies.

More often than not, gas accreting onto a black hole is not able to radiate efficiently (*14*). Dissipated energy is retained as heat, generating pressure that inflates the flow into a geometrically thick disk or torus. There is a reciprocal relation between the temperature of the gas and the distance from the black hole in units of R_S. When this ratio is smaller than 10^3, the gas temperature exceeds 10^9 K and electron thermal velocities approach the speed of light. Under these conditions, electrons lose energy rapidly. If they are continually resupplied with energy by the ions, which do not radiate efficiently, they will quickly drain away the heat in the accretion flow, which will then settle into a thin disk. This implies that the gas in a nonradiative accretion flow must be characterized by two temperatures; that is, the electron component is much colder than the ion component (*35, 36*). Even where Coulomb collisions are unable to keep electrons and ions in equipartition, plasma instabilities have the potential to transfer energy between the two species. The fact that this does not seem to happen may indicate that the magnetic energy density remains small relative to the thermal pressure (*37*).

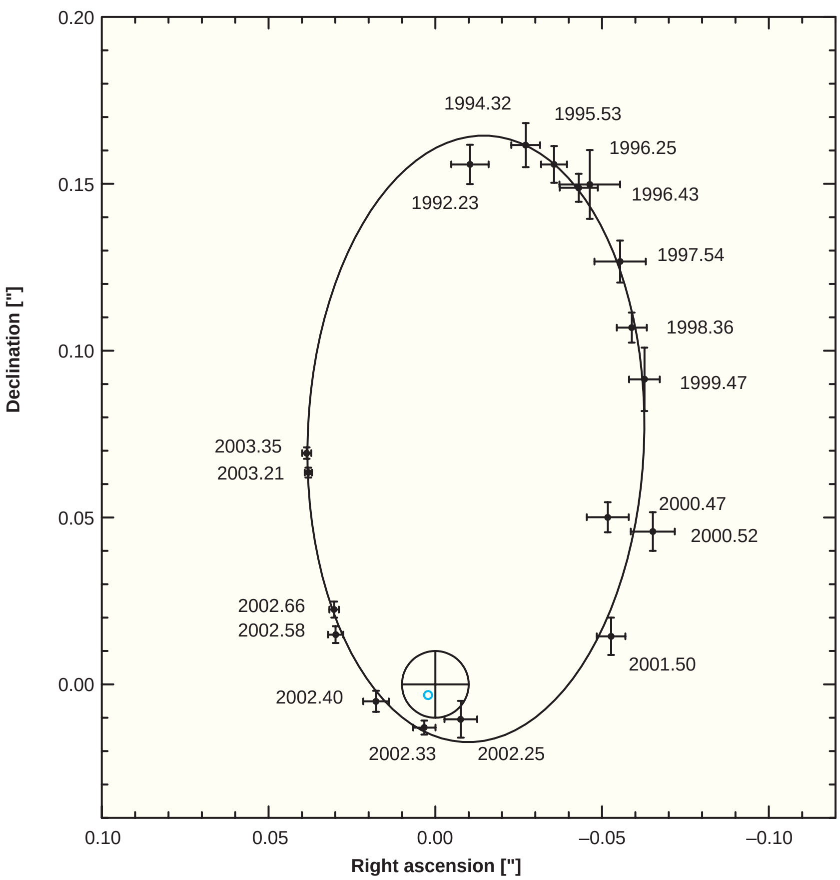


Fig. 2. Orbit of the star S2 around Sgr A* [figure 1 of (*17*)]. The continuous curve shows the projected best fit Keplerian orbit, which has a period of 15.6 years and comes within 2000 Schwarzschild radii of the position of Sgr A* (shown as the large cross within a circle). The small blue open circle marks the focus of the elliptical orbit. Ghez *et al*. (*16*) report the orbits of several additional stars, including one that comes within 1000 R_S of the putative black hole. Dates within years are shown as decimal fractions of years.

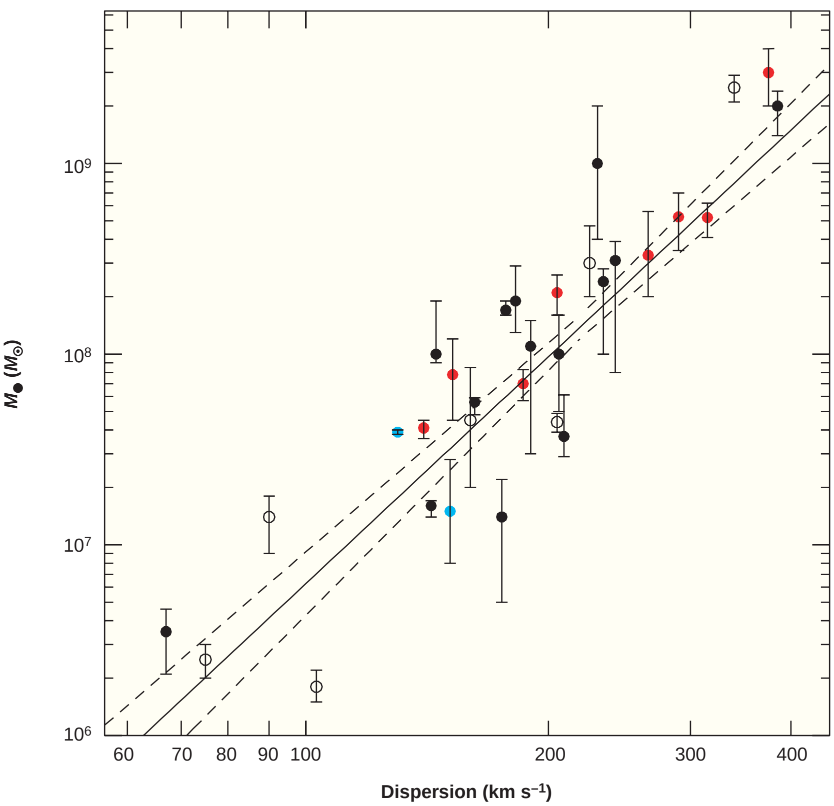


Fig. 3. Black hole masses plotted against bulge velocity dispersions of the host galaxy [figure 7 of (*22*); reproduced by permission of the American Astronomical Society]. Mass measurements are based on stellar (open and solid black circles), gas (solid red circles), and maser (solid green circles) kinematics. The solid line shows the best-fit $M - \sigma$ correlation; the dashed lines show one standard deviation on either side. Open and solid circles denote data from two different research groups.

If the energy liberated by accretion is not radiated away, where does it go? One possibility is that it is advected into the black hole (*38*). However, this leads to serious stability problems and is probably untenable. The reason is that much of the energy liberated close to the black hole is transferred, by the torque, to material farther out. This leaves the distant material with more than enough energy to escape from the black hole's gravitational field. The likely result is that only a small fraction of the gas that comes under the hole's gravitational influence, on the order of the ratio of R_S to the accretion radius (where the free-fall speed first equals the sound speed in the gas), is actually accreted. The rest is probably expelled by the outward flux of energy before it comes near the horizon (*39–41*) (Fig. 4). This tendency of black holes to reject all but a tiny fraction (typically, 10^{-5} or smaller) of the matter supplied to them can explain why supermassive black holes are often so underluminous, despite their gas-rich environments. The outward energy transport converts a steep density profile ($\propto r^{-3/2}$) into a much shallower one ($\propto r^{-1/2}$), diminishing the emissivity of the gas close to the horizon (*42*).

Relativistic jets are the most striking manifestations of outflow from the vicinity of black holes. Observations show that jets are accelerated and collimated close to the black hole (*43*), probably by magnetic fields (*7*). In AGN they reach speeds as high as 90 to 99.9% of the speed of light [corresponding to a range of Lorentz factors of ~2 to 20 (*44*)] and, in some cases, retain a highly relativistic velocity and tight collimation out to enormous distances from the black hole.

Jets seem to be a generic way for accretion disks to rid themselves of excess energy and angular momentum; they also appear in subrelativistic systems such as protostars. However, the energy source for black hole jets need not be limited to the accretion flow. Magnetic fields, supported by currents in the external gas, can extract energy from the spin of the black hole via the Blandford-Znajek (BZ) effect (*45*). Because the spin energy of a black hole resides in the spacetime outside the horizon (mainly in the region known as the ergosphere, where all trajectories are dragged in the direction of spin), no physical laws are violated when a black hole is spun down. In theory, up to 29% of a black hole's rest mass energy can be liberated in this way. If some of the magnetic field lines that couple to the black hole spin also thread the accretion disk, the BZ effect could also contribute to the disk's luminosity (*34*).

The relative contributions of accretion power and the BZ effect to jet production are uncertain. Jets are found in only about 10% of AGN, including objects that appear to have accretion flows similar to systems without jets. Curiously, AGN with powerful jets are almost exclusively associated with elliptical galaxies, which suggests that somehow the black hole knows about galactic structure on scales eight orders of magnitude larger than the event horizon. According to the spin hypothesis, the connection arises because black holes in ellipticals are spinning at close to the maximal rate as a result of their history of mergers (*46*), and are thus capable of generating large jet powers via the BZ effect. However, other studies suggest that black hole spin might not depend so sensitively on galactic environment (*47, 48*). Other factors, such as the magnetic topology or outer boundary conditions of the accretion flow, may be as important (or more important) in determining whether jets form.

In this connection, the discovery of jets from a class of x-ray binaries called microquasars is noteworthy (*49*). Jets from microquasars are not produced continuously, but emerge after outbursts that may represent the sudden draining of the inner accretion disk. Because dynamical time scales close to the event horizon are proportional to the black hole mass, events that last thousands to hundreds of thousands of years in AGN may have analogs lasting minutes to days in microquasars, giving us a time-lapse view of evolving phenomena that would otherwise escape detection. As discussed below, the speeded-up phenomenology of microquasars may allow us to measure the spins of their black holes, providing a test for the spin hypothesis.

Are They Really Black Holes?

Despite the overwhelming circumstantial evidence for black holes, the measurements discussed so far do not establish that the dark masses and compact objects we detect are the black holes whose properties are predicted so precisely by general relativity. Even the enormous release of energy during accretion will occur in any gravitational potential well of comparable depth—for example, that of a neutron star. To really confirm the existence of black holes and to test general relativity in the strong gravity limit, we must devise diagnostics sensitive to the curvature of spacetime near the horizon. Such measurements are now being made with varying degrees of success. Note that the existence of an event horizon is the only truly distinctive feature of a black hole. Neutron stars may have an innermost stable circular orbit if they are sufficiently compact (*6*). Likewise, the dragging of inertial frames occurs around any body with angular momentum, although its effects should be most pronounced close to a rapidly spinning black hole. All three phenomena, however, are consequences of the curvature of spacetime according to general relativity.

Evidence for the event horizon has been surprisingly difficult to establish observationally. One of the most frustrating aspects of

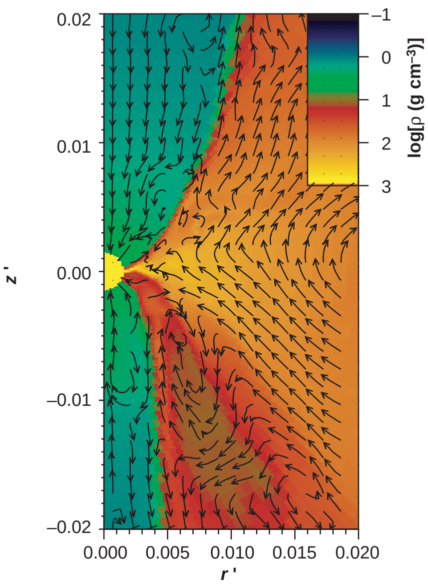


Fig. 4. Snapshot of logarithmic density overplotted by direction of the poloidal velocity, from a two-dimensional simulation of magnetohydrodynamic flow onto a black hole [figure 8b of (*41*)]. Matter is supplied continuously at a scaled radius of 1.2 (not shown), along with a weak radial magnetic field. Conditions near the black hole are modeled by a "pseudo-Newtonian" gravitational potential that mimics the effects of general relativity. Magnetic stresses generated inside the rotating torus (red-orange) block low-angular-momentum material (green) from accreting along the rotational axis, and expel most of the rotating material as well. As a result, only 1% of the matter supplied at r = 1.2 is swallowed by the black hole.

studying x-ray binaries has been the difficulty in distinguishing neutron stars from black holes by means other than mass estimates. Neutron stars have surfaces that halt the inflow of matter, whereas black holes have horizons through which matter passes freely; hence, accreting neutron stars should exhibit some extra emissivity, and perhaps a distinctive spectral signature, from matter impacting the surface. No clear spectral discriminants have been identified, but there are differences between the x-ray luminosities of quiescent x-ray novae—essentially microquasars—containing black hole and neutron stars (*50*). For the comparison to be meaningful, the sample must be carefully controlled so that the binaries have similar mass-transfer rates, which are thought to correlate with the orbital period. The results are suggestive but not definitive. The mass-transfer rates refer to the mass supplied to the outer accretion flow, not the accretion rate actually reaching the compact object. Because the inner accretion flows in quiescence are radiatively inefficient, only a small fraction of the supplied mass presumably reaches the compact object (*39*). It is not known whether there are systematic differences in the accreted fraction (attributable, e.g., to the different masses of the primaries or the different radiation environments). Moreover, residual accretion energy left over from outbursts would be stored in the neutron star's crust and emitted during quiescence (*51*), so it is not even clear that the observed emission reflects the real-time accretion rate.

Prospects are much better for detecting phenomena associated with the ISCO. The ISCO, which depends on both the mass and spin of the black hole, sets a characteristic inner radius for an accretion disk. X-ray observations of extremely broad spectral lines from partially ionized iron have provided direct evidence for disk-like flow close to the ISCO (*52*). These lines are thought to arise from fluorescence of relatively cool, optically thick gas exposed to hard x-rays produced in an optically thin corona. If the fluorescing gas forms the inner part of an accretion disk orbiting a black hole, then the line profile should display a relatively narrow blue wing boosted in intensity by the radial Doppler shift, and a broad red wing shaped by the combination of gravitational redshift and transverse Doppler shift. The best studied case, the Seyfert galaxy MCG–6-30-15, shows these features (Fig. 5), and its rapid variability confirms that the line is produced close to the horizon. The line profile can be used to deduce the spin of the black hole, although the fit may be nonunique because the structure of the corona is weakly constrained and emission could arise from inside the ISCO (*53*). In a couple of observed cases, iron line spectra may be revealing the dissipation of spin energy extracted from a rapidly rotating black hole and deposited in the innermost regions of the disk. Both MCG–6-30-15 (*54*) and the microquasar XTE J1650-500 (*55*) show such extreme redshifted emission that they are difficult to reconcile with any model in which the power in the line derives from gravitational binding energy liberated by the accretion flow.

In addition to setting a length scale, the ISCO sets a variety of time scales, including an orbital time, vertical and radial oscillation time scales for perturbations about nearly circular orbits, and a precessional time scale associated with the dragging of inertial frames by the spin of the black hole. Because these time scales are properties of the black hole's spacetime rather than the gas dynamics of the accretion flow, they should define specific frequencies that are insensitive to fluctuations in the luminosity or spectrum. Moreover, these frequencies should be among the highest associated with an accreting black hole, ranging from μHz for the most massive AGN to several hundred Hz for x-ray binaries. At least five microquasars show quasi-periodic oscillations (QPOs) at stable, high frequencies (*56*). Three of these show pairs of QPOs with simple frequency ratios, suggesting resonance effects. Although the mechanism that creates the modulations is unknown, attention has been drawn to diskoseismic modes of accretion disks. According to general relativity, the radial oscillation frequency has a maximum outside the ISCO and declines both inward (vanishing at the ISCO, which is why circular orbits become unstable there) and outward (where it approaches the Keplerian orbital frequency). Consequently, the inner part of an accretion disk around a black hole (or a sufficiently compact neutron star with a weak magnetic field) can behave like a resonant cavity, capable of trapping and amplifying wave modes that resonate with the various characteristic frequencies (*57*). The relevant modes, their amplification mechanisms, and their spectral signatures still remain to be identified. Once this is done, we should be able to measure black hole spins as reliably as we can measure their masses and test for other predicted features of spacetime curvature as well.

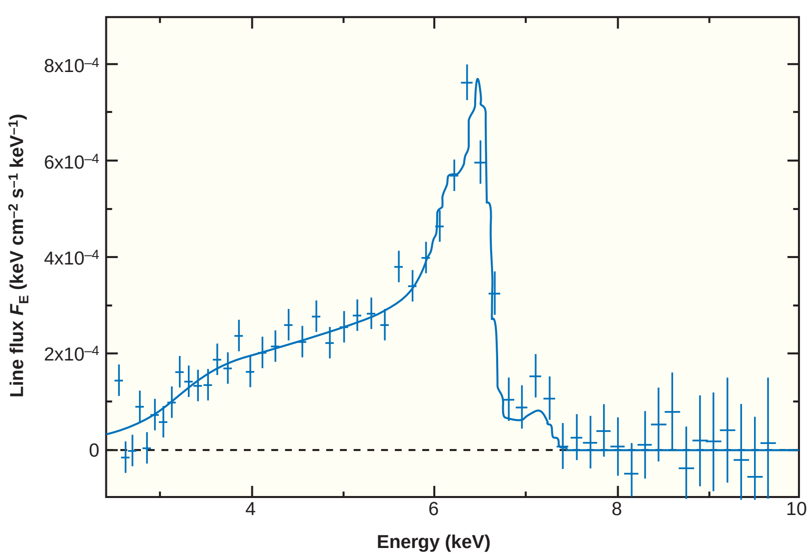


Fig. 5. Spectrum of the broad iron Kα line in MCG–6-30-15 taken with the XMM-Newton satellite [figure 3 of (*52*)], showing the combined effects of spectral distortions described in Fig. 1B.

Prospects

The evidence for black holes has firmed up substantially during the past 5 years, and we have every reason to expect the pace of discovery to accelerate. High-resolution x-ray spectroscopy and timing measurements will continue with existing satellites and will improve with the next generation of x-ray observatories, notably Constellation-X. Long-duration observations will test whether the high-frequency oscillations observed in microquasars also exist, at scaled-down frequencies, in their more massive counterparts. Understanding the demographics of black hole spin as well as mass will give us a much clearer idea of how black holes formed. A longer term but realistic goal is the direct imaging by x-ray interferometry of the accretion disk around a black hole (Fig. 1B).

Whereas x-ray measurements probe the stationary spacetimes of black holes, gravitational wave detectors will enable us to study the dynamics of spacetime as black holes form and merge. A future generation of the Laser Interferometer Gravitational-Wave Observatory (LIGO) may reveal which dying stars form black holes promptly, which undergo delayed collapse, and whether gamma-ray bursts really represent the birth of stellar-mass black holes. The Laser Interferometer Space Antenna (LISA), tuned to the slower pace of supermassive black holes, should detect stellar-mass objects falling into supermassive black holes as well as the mergers of supermassive black hole binaries (*58*). Even before LISA flies, we may see the clear signal of a star being torn apart and swallowed by a supermassive black hole (*59*).

Numerical simulations are finally reaching the point (limited mainly by computer speed) at which we can perform fully three-dimensional, magnetohydrodynamic simulations of accretion flows onto black holes. General relativistic codes are being developed and tested. We look forward to codes that can handle the microphysics of the gas as well, which will enable us to address a variety of questions: How is magnetic energy dissipated during accretion? Are two-temperature flows possible? How much matter and energy is ejected from accretion flows? What propels jets? We also anticipate progress in simulating the effects of black holes on their large-scale environments. Incorporated into cosmological simulations, these feedback calculations may clarify the origin of the "$M - \sigma$" relation and the role of black holes in galaxy formation.

Physicists will continue to study the interiors of black holes intensively (and theoretically) for clues to the fundamental structure of matter, the quantum nature of spacetime, and the possible existence of extra dimensions. For the public, black holes will retain their metaphorical implications of disappearance and mystery. Astrophysicists, on the other hand, have recently appreciated how commonplace black holes are. As we learn more about their formation and how they interact with their environments, we will understand their roles in shaping the formation and evolution of the galaxies we see around us.

References and Notes

1. K. S. Thorne, *Black Holes and Time Warps: Einstein's Outrageous Legacy* (Norton, New York, 1994).
2. J. B. Hartle, *Gravity: An Introduction to Einstein's General Relativity* (Addison-Wesley, San Francisco, 2003).
3. J. Kormendy, K. Gebhardt, in *Proceedings of the 20th Texas Symposium on Relativistic Astrophysics,* J. C. Wheeler, H. Martel, Eds. (American Institute of Physics, Melville, NY, 2001), pp. 363–381 (available at http://xxx.lanl.gov/abs/astro-ph/0105230).
4. E. Maoz, *Astrophys. J.* **494**, L181 (1998).
5. P. Charles, in *Black Holes in Binaries and Galactic Nuclei,* L. Kaper, E. P. J. van den Heuvel, P. A. Woudt, Eds. (Springer, Berlin, 2001), pp. 27–38.
6. J. M. Lattimer, M. Prakash, *Astrophys. J.* **550**, 426 (2001).
7. D. L. Meier, S. Koide, Y. Uchida, *Science* **291**, 84 (2001).
8. M. C. Begelman, in *Coevolution of Black Holes and Galaxies,* L. C. Ho, Ed. (Cambridge Univ. Press, Cambridge, in press) (available at http://xxx.lanl.gov/abs/astro-ph/0303040).
9. T. Piran, *Phys. Rep.* **314**, 575 (1999).
10. D. P. Bennett *et al.*, *Astrophys. J.* **579**, 639 (2002).
11. S. Mao *et al.*, *Mon. Not. R. Astron. Soc.* **329**, 349 (2002).
12. C. L. Fryer, *Astrophys. J.* **522**, 413 (1999).
13. J. H. Krolik, *Active Galactic Nuclei: From the Central Black Hole to the Galactic Environment* (Princeton Univ. Press, Princeton, NJ, 1999).
14. T. Di Matteo, C. L. Carilli, A. C. Fabian, *Astrophys. J.* **547**, 731 (2001).
15. R. Schödel *et al.*, *Nature* **419**, 694 (2002).
16. A. M. Ghez *et al.*, *Astrophys. J.*, in press (available at http://xxx.lanl.gov/abs/astro-ph/0302299).
17. F. Eisenhauer *et al.*, in preparation.
18. F. K. Baganoff *et al.*, *Nature* **413**, 45 (2001).
19. M. Miyoshi *et al.*, *Nature* **373**, 127 (1995).
20. L. Ferrarese, D. Merritt, *Astrophys. J.* **539**, L9 (2000).
21. K. Gebhardt *et al.*, *Astrophys. J.* **539**, L13 (2000).
22. S. Tremaine *et al.*, *Astrophys. J.* **574**, 740 (2002).
23. Q. Yu, S. Tremaine, *Mon. Not. R. Astron. Soc.* **335**, 965 (2002).
24. R. D. Blandford, in *Galaxy Dynamics*, D. R. Merritt, M. Valluri, J. A. Sellwood, Eds. (Astronomical Society of the Pacific, San Francisco, 1999), pp. 87–95 (available at http://xxx.lanl.gov/abs/astro-ph/9906025).
25. A. C. Fabian, *Mon. Not. R. Astron. Soc.* **308**, L39 (1999).
26. A. El-Zant, I. Shlosman, M. C. Begelman, J. Frank, *Astrophys. J.*, in press (available at http://xxx.lanl.gov/abs/astro-ph/0301338).
27. R. P. van der Marel, in *Coevolution of Black Holes and Galaxies*, L. C. Ho, Ed. (Cambridge Univ. Press, Cambridge, in press) (available at http://xxx.lanl.gov/abs/astro-ph/0302101).
28. E. J. M. Colbert, R. F. Mushotzky, *Astrophys. J.* **519**, 89 (1999).
29. J. M. Miller, G. Fabbiano, M. C. Miller, A. C. Fabian, *Astrophys. J.* **585**, L37 (2003).
30. A. R. King *et al.*, *Astrophys. J.* **552**, L109 (2001).
31. M. C. Begelman, *Astrophys. J.* **568**, L97 (2002).
32. S. A. Balbus, J. F. Hawley, *Rev. Mod. Phys.* **70**, 1 (1998).
33. N. I. Shakura, R. A. Sunyaev, *Astron. Astrophys.* **24**, 337 (1973).
34. E. Agol, J. H. Krolik, *Astrophys. J.* **528**, 162 (2000).
35. S. Ichimaru, *Astrophys. J.* **214**, 840 (1977).
36. M. J. Rees, M. C. Begelman, R. D. Blandford, E. S. Phinney, *Nature* **295**, 17 (1982).
37. E. Quataert, A. Gruzinov, *Astrophys. J.* **520**, 248 (1999).
38. R. Narayan, I. Yi, *Astrophys. J.* **452**, 710 (1995).
39. R. D. Blandford, M. C. Begelman, *Mon. Not. R. Astron. Soc.* **303**, L1 (1999).
40. J. F. Hawley, S. A. Balbus, *Astrophys. J.* **573**, 738 (2002).
41. D. Proga, M. C. Begelman, *Astrophys. J.*, in press (available at http://xxx.lanl.gov/abs/astro-ph/0303093).
42. R. Narayan, I. V. Igumenshchev, M. A. Abramowicz, *Astrophys. J.* **539**, 798 (2000).
43. W. Junor, J. A. Biretta, *Astron. J.* **109**, 500 (1995).
44. The Lorentz factor is a convenient way of characterizing velocities very close to the speed of light. It is defined as $\Gamma = (1 - v^2/c^2)^{-1/2}$, where v is the velocity, and $\Gamma \rightarrow \infty$ as $v \rightarrow c$. The Lorentz factor plays a fundamental role in relativistic calculations; for example, the kinetic energy and effective mass of a rapidly moving body are proportional to Γ, whereas time in a rapidly moving frame of reference is slowed down, or dilated, by a factor Γ^{-1}.
45. R. D. Blandford, R. L. Znajek, *Mon. Not. R. Astron. Soc.* **179**, 433 (1977).
46. A. S. Wilson, E. J. M. Colbert, *Astrophys. J.* **438**, 62 (1995).
47. R. Moderski, M. Sikora, J.-P. Lasota, *Mon. Not. R. Astron. Soc.* **301**, 14 (1998).
48. S. A. Hughes, R. D. Blandford, *Astrophys. J.* **585**, 101 (2003).
49. I. F. Mirabel, L. F. Rodríguez, *Annu. Rev. Astron. Astrophys.* **37**, 409 (1999).
50. M. R. Garcia *et al.*, *Astrophys. J.* **553**, L47 (2001).
51. E. F. Brown, L. Bildsten, R. E. Rutledge, *Astrophys. J.* **504**, L95 (1998).
52. A. C. Fabian *et al.*, *Mon. Not. R. Astron. Soc.* **335**, L1 (2002).
53. C. S. Reynolds, M. C. Begelman, *Astrophys. J.* **488**, 109 (1997).
54. J. Wilms *et al.*, *Mon. Not. R. Astron. Soc.* **328**, L27 (2001).
55. J. M. Miller *et al.*, *Astrophys. J.* **570**, L69 (2002).
56. R. A. Remillard, M. P. Muno, J. E. McClintock, J. A. Orosz, *Astrophys. J.* **580**, 1030 (2002).
57. R. V. Wagoner, A. S. Silbergleit, M. Ortega-Rodriguez, *Astrophys. J.* **559**, L25 (2001).
58. M. C. Begelman, R. D. Blandford, M. J. Rees, *Nature* **287**, 307 (1980).
59. M. J. Rees, *Nature* **333**, 523 (1988).

60. Supported by NSF grant AST-9876887.

66. P. Mészáros, M. J. Rees, *Astrophys. J.* **476**, 231 (1997).
67. B. Zhang, S.Kobayashi, P. Mészáros, *Astrophys. J.*, in press (e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302525).
68. J. Miralda-Escudé, *Astrophys. J.* **501**, 15 (1998).
69. D. Q. Lamb, D. Reichart, *Astrophys. J.* **536**, 1 (2000).
70. V. Bromm, A. Loeb, *Astrophys. J.* **575**, 111 (2002).
71. S. P. Oh, *Mon. Not. R. Astron. Soc.* **336**, 1021 (2002).
72. S. R. Furlanetto, A. Loeb, *Astrophys. J.* **579**, 1 (2002).
73. A. Kogut *et al.*, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302213.
74. The errors are 1σ and include the best estimate of the WMAP team of systematic uncertainties associated with foreground Galactic emission; see table 2 of (*73*).
75. J. Miralda-Escudé, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0211071.
76. S. Whyithe, A. Loeb, *Astrophys. J. Lett.*, in press (e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302297).
77. Z. Haiman, G. Holder, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302403.
78. G. Holder, Z. Haiman, M. Kaplinghat, L. Knox, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302404.
79. B. Ciardi, A. Ferrara, S. D. M. White, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0302451.
80. R. S. Somerville, M. Livio, e-Print available at http://xxx.lanl.gov/abs/ astro-ph/0303017.
81. R. S. Somerville, J. S. Bullock, M. Livio, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0303481.
82. A. Sokasian, T. Abel, L.Hernquist, V. Springel, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0303098.
83. R. Cen, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0303236.
84. W. A. Chiu, X. Fan, J. P. Ostriker, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0304234.
85. The maximum theoretical upper limit for the number of ionizing photons that can be emitted by a star for every baryon it contains is obtained by dividing the fusion energy per baryon released by fusion to helium, 7 MeV, by the average energy of an ionizing photon, $\sim$20 eV, which gives $\sim 3 \times 10^5$. Massive stars with no heavy elements can fuse almost all their hydrogen content over their lifetime and are hot enough to emit most of their radiation as ionizing photons (*86–88*).
86. J. Tumlinson, J. M. Shull, *Astrophys. J.* **528**, L65 (2000).
87. V. Bromm, R. P. Kudritzki, A. Loeb, *Astrophys. J.* **552**, 464 (2001).
88. D. Schaerer, *Astron. Astrophys.* **382**, 28 (2002).
89. A. Venkatesan, J. Tumlinson, J. M. Shull, *Astrophys. J.* **584**, 621 (2003).
90. D. Scott, M. J. Rees, *Mon. Not. R. Astron. Soc.* **247**, 510 (1990).
91. S. A. Wouthuysen, *Astron. J.* **57**, 31 (1952).
92. G. B. Field, *Astrophys. J.* **129**, 551 (1959).
93. P. Madau, A. Meiksin, M. J. Rees, *Astrophys. J.* **475**, 492 (1997).
94. X. Chen, J. Miralda-Escudé, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0303395.
95. P. Tozzi, P.Madau, A. Meiksin, M. J. Rees, *Astrophys. J.* **528**, 597 (2000).
96. I. T. Iliev, P. R. Shapiro, A. Ferrara, H. Martel, *Astrophys. J.* **572**, 123 (2002).
97. C. L. Carilli, N. Y. Gnedin, F. Owen, *Astrophys. J.* **577**, 22 (2002).
98. S. R. Furlanetto, A. Loeb, *Astrophys. J.* **588**, 18 (2003).
99. B. Ciardi, P. Madau, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0303249.
100. U.-L. Pen, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0305387.
101. At present the Giant Metrewave Radio Telescope (www.gmrt.ncra.tifr.res.in) in India is already searching for high-*z* 21-cm signals, and more sensitive observatories being designed now are the Square Kilometer Array (www.skatelescope.org) and the Low Frequency Array (www.lofar.org).
102. W. H. Press, P. Schechter, *Astrophys. J.* **187**, 425 (1974).
103. J. R. Bond, S. Cole, G. Efstathiou, N. Kaiser, *Astrophys. J.* **379**, 440 (1991).
104. R. J. Bower, *Mon. Not. R. Astron. Soc.* **248**, 332 (1991).
105. C. Lacey, S. Cole, *Mon. Not. R. Astron. Soc.* **262**, 627 (1993).

106. I thank X. Fan and N. Gnedin for their permission to reproduce and their help in providing figures from their papers and P. Sieber for suggesting a good way to start this article. I also thank T. Abel, A. Loeb, M. Rees, and my referees for their comments.

REVIEW

New Light on Dark Matter

Jeremiah P. Ostriker[1] and Paul Steinhardt[2]

Dark matter, proposed decades ago as a speculative component of the universe, is now known to be the vital ingredient in the cosmos: six times more abundant than ordinary matter, one-quarter of the total energy density, and the component that has controlled the growth of structure in the universe. Its nature remains a mystery, but assuming that it is composed of weakly interacting subatomic particles, is consistent with large-scale cosmic structure. However, recent analyses of structure on galactic and subgalactic scales have suggested discrepancies and stimulated numerous alternative proposals. We discuss how studies of the density, demography, history, and environment of smaller-scale structures may distinguish among these possibilities and shed new light on the nature of dark matter.

The dark side of the universe first became evident about 65 years ago when Fritz Zwicky (*1*) noticed that the speed of galaxies in large clusters is much too great to keep them gravitationally bound together unless they weigh over 100 times more than one would estimate on the basis of the number of stars in the cluster. Decades of investigation confirmed his analysis (*2–5*), and by the 1980s, the evidence for dark matter with an abundance of about 20% of the total energy density of the universe was accepted, although the nature of the dark matter remained a mystery.

After the introduction of inflationary theory (*6*), many cosmologists became convinced that the universe must be flat and that the total energy density must equal the value (termed the critical value) that distinguishes a positively curved, closed universe from a negatively curved, open universe. Cosmologists became attracted to the beguiling simplicity of a universe in which virtually all of the energy density consists of some form of matter, about 4% being ordinary matter and 96% dark matter. In fact, observational studies were never really compliant with this vision. Although there was a wide dispersion in total mass density estimates, there never developed any convincing evidence that there was sufficient matter to reach the critical value. The discrepancy between observation and the favored theoretical model became increasingly sharp.

Dark energy came to the rescue when it was realized that there was not sufficient matter to explain the structure and nature of the universe (*7*). The only thing dark energy has in common with dark matter is that both components neither emit nor absorb light. On a microscopic scale, they are composed of different constituents. Most important, dark matter, like ordinary matter, is gravitationally self-attractive and clusters with ordinary matter to form galaxies. Dark energy is gravitationally self-repulsive and remains nearly uniformly spread throughout the universe. Hence, a census of the energy contained in galaxies would miss most the dark energy. So, by positing the existence of a dark energy component, it became possible to account for the 70 to 80% discrepancy between the measured mass density and the critical energy density predicted by inflation (*8–11*). Then, two independent groups (*12*, *13*) found evidence of the accelerated expansion of the universe from observations of supernovae, and the model with a dominant dark energy component, as illustrated in Fig. 1, became the concordance model of cosmology. The existence of dark energy has recently been independently confirmed by observations by the Wilkinson Microwave Anisotrope Probe [WMAP (*14*)] and has become accepted as an essential ingredient of the standard model (*15*).

Dark energy has changed our view of the role of dark matter in the universe. According to

[1]Department of Astrophysical Sciences, [2]Department of Physics, Princeton University, Princeton, NJ 08544, USA.

Einstein's general theory of relativity, in a universe composed only of matter, it is the mass density that determines the geometry, the history, and the future of the universe. With the addition of dark energy, the story is different. First, what determines the geometry of the universe is whether the total energy density equals the critical value, where now we add to the mass contribution (identifying its energy according to $E = mc^2$) the dark energy contribution. Second, the period of matter domination has given way to dark energy domination. So, the important role of dark matter is in the past, when it was the dominant contribution to the energy density; roughly the first few billion years. Our future is determined by the nature of the dark energy, which is sufficient to cause the current expansion of the universe to accelerate, and the acceleration will continue unless the dark energy should decay or change its equation of state.

We have neglected one very important subplot up to this point: dark matter as the agent producing the growth of cosmic structure. We would not exist today were it not for dark matter, which played a crucial role in the formation of the present structure in the universe. Without dark matter, the universe would have remained too uniform to form the galaxies, stars, and planets. The universe, although nearly homogeneous and isotropic on its largest scales, shows a bewildering variety of structures on smaller scales: Stars, galaxies, clusters of galaxies, voids, and great walls of galaxies have been found. The only known force capable of moving matter on such large scales is Newton's gravity. And because, in a smooth and uniform medium, there will be no irregularities to produce gravitational forces, all structures must have been seeded by small fluctuations imprinted on the universe at very early times. These fluctuations should leave a signature on the cosmic background radiation (CBR) left over from the Big Bang. Ordinary matter could not produce fluctuations to create any substantial structures without leaving a signal bigger than what was observed in the CBR, because it remained tightly coupled to radiation, preventing it from clustering, until recent epochs.

On the other hand, dark matter, which is not coupled to photons, would permit tiny fluctuations (consistent with the CBR observations) to grow for a long, long time before the ordinary matter decoupled from radiation. Then, the ordinary matter would be rapidly drawn to the dense clumps of dark matter and form the observed structures. There would still need to be initial fluctuations, but their amplitude could be substantially smaller than otherwise. The required material was called cold dark matter, because it consisted of nonrelativistic particles that were assumed to contain no internal thermal motions (that is, they were cold).

A final important ingredient in the standard paradigm must be mentioned before we can begin to assess the validity of the picture. The initial spectrum of perturbations (the ratio of long waves to short waves) must be specified in order to predict the gravitational effects of these waves. The initial density fluctuations were scale-invariant. That is, if we decomposed the energy distribution into a sum of sinusoidal waves of varying wavelengths, the wave amplitudes of the waves were the same for all wavelengths. One of the great triumphs of the inflationary scenario (*16–20*) is that it provided a well-motivated dynamical mechanism for producing a nearly scale-invariant (defined by spectral index $n = 1$) spectrum. This prediction has now been confirmed by the WMAP, which found $n = 0.99 \pm 0.04$ (*21*).

But we cannot claim to understand the the universe if we do not know the nature of dark matter. Two kinds of dark matter are already known, neutrinos and black holes (*22*), but they are thought to make minor contributions to the total. The majority component remains unknown. Here we explore these issues: the possible candidates, their implications for structure formation, and how we might use a combination of particle detectors and astronomical observations to resolve the nature of dark matter.

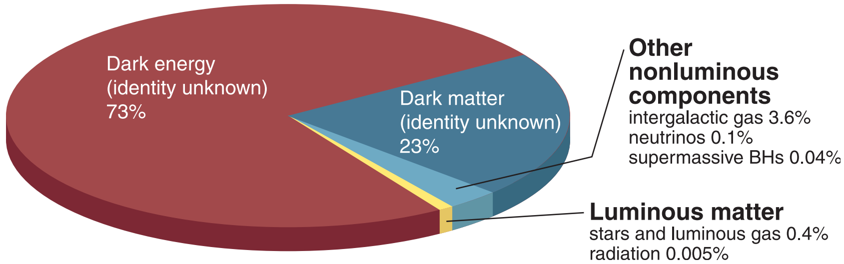


Fig. 1. The luminous (light-emitting) components of the universe only comprise about 0.4% of the total energy. The remaining components are dark. Of those, roughly 3.7% are identified: cold gas and dust, neutrinos, and black holes. Nearly 23% is dark matter, and the overwhelming majority is some type of gravitationally self-repulsive dark energy.

The Favored Candidates for Dark Matter

For over a decade, the favored candidates for dark matter have been hypothetical elementary particles that are long-lived, cold, and collisionless. Long-lived means the lifetime must be comparable to or greater than the present age of the universe, about 14 billion years. Cold means that the particles are nonrelativistic at the onset of the matter-dominated epoch, so that they are immediately able to cluster gravitationally. Because clustering occurs on length scales smaller than the Hubble horizon (the age of the universe multiplied by the speed of light), and the Hubble horizon was much smaller during the era of matter domination than today, the first objects to form—clumps or halos of dark matter—were much smaller and less massive than the Milky Way. As the universe expanded and the Hubble horizon grew, many of these first small halos merged to form larger-scale structures, which later merged to form yet larger-scale structures. The result is a hierarchy of structure ranging over many orders of magnitude in volume and mass, which is qualitatively in accordance with what is observed. In contrast, hot relativistic particles, such as light massive neutrinos, would be moving too fast during the time of matter domination to cluster gravitationally, and would result in a distribution of structure that is inconsistent with what is observed. Hence, light neutrinos must be a negligible component of the dark matter mass density, a conclusion supported by measurements of the neutrino mass in underground solar neutrino experiments. Collisionless means that the interaction cross-section between dark matter particles (and between dark matter and ordinary matter) is so small as to be negligible for densities found in dark matter halos. The particles are only gravitationally bound to one another and travel unimpeded in orbits in the halos with a broad spectrum of eccentricities.

Cold collisionless dark matter (CCDM) has been favored for several reasons. First, numerical simulations of structure formation with CCDM agree with most observations of structure. Second, for a special subclass known as WIMPs (weakly interacting massive particles), there is a natural explanation for why they have the requisite abundance. If particles interact through the weak force, then they were in thermal equilibrium in the first trillionths of a second after the Big Bang, when the density and temperature were high. Then they fell out of equilibrium, with a concentration that is predicted from their annihilation cross-section. For a weak force cross-section, the expected mass density today spans a range that includes 20 to 30% of the total energy density of the universe, as observed. A third reason for favoring CCDM is that there are specific appealing candidates for the particles in models.

One candidate is the neutralino, a particle that arises in models with supersymmetry. Supersymmetry, a fundamental aspect of supergravity and superstring theories, requires a (yet unobserved) boson partner particle for every known fermion and a fermion partner particle for every known boson. If supersymmetry were extant today, the partners would have the same mass. But because supersymmetry would have been spontaneously broken at high temperatures in the early universe, today the masses are different. Also, most supersymmetric partners are unstable and decayed soon after the breaking of symmetry. However, there is a lightest partner (with mass on the order of 100 GeV) that is prevented by its symmetries from decaying. In the simplest models, these particles are electrically neutral and weakly interacting—ideal candidates for WIMPs. If the dark matter consists of neutralinos, then underground detectors can detect their passage through Earth as the planet travels around the Sun and through the dark matter in the solar neighborhood. However, it is important to note that detection alone does not necessarily mean that dark matter consists primarily of WIMPS. The current experiments cannot determine whether WIMPS are a majority or, like neutrinos, a small minority of the dark matter.

Another appealing candidate is the axion, a very light neutral particle (with mass on the order of 1 μeV) that is important in suppressing strong CP violation in unified theories. The axion interacts through such a tiny force that it is never in thermal equilibrium, so the explanation for its abundance is not as simple. It immediately forms a cold Bose condensate that permeates the universe. Axion detectors have been constructed and the search for them is under way.

Cracks in the Foundation

Because the standard model, combined with CCDM, is mathematically quite specific (even if some of the parameters that enter into it are imprecisely known), it can be tested at many different scales. The largest scales (thousands of megaparsecs) are seen in the CBR. CBR measurements show the primordial distribution of energy and matter when their distribution was nearly uniform and there was no structure. Next come measurements of the large-scale structure seen in the distribution of galaxies ranging from several Mpc to nearly 1000 Mpc. Over these scales, observation and theory are consistent, inspiring great confidence in the overall picture.

However, on smaller scales, from 1 Mpc down to the scale of galaxies, kiloparsecs, and below, there is inconsistency. These apparent disagreements began to surface several years ago (*23–25*), and no consensus has emerged as to whether they represent real problems. For the most part, theorists believe that, if there is a problem, it is much more likely to be due to our specific assumptions about the nature of dark matter than to a problem with the global picture given by the standard model. That there should be more uncertainty about smaller objects that are relatively closer may seem puzzling at first, but there are natural explanations. First, on large scales gravity is dominant, so an understanding of the predictions involves only computations based on Newton's and Einstein's laws of gravity. On smaller scales, the complex hydrodynamical interactions of hot dense matter must be included. Second, the fluctuations on large scales are small and we have accurate methods of computing such quantities. But on the scales of galaxies, the physical interactions of ordinary matter and radiation are more complex. The principal purported problems found on smaller scales are as follows: Substructure—small halos and galaxies orbiting within larger units—may not be as common as expected on the basis of numerical simulations of CCDM. The number of halos expected varies roughly as the inverse of the mass, so many more dwarf galaxy systems should have been observed. The lensing effect of small halos should be evident from the distribution of brightnesses of multiple images of a given galaxy, but the current evidence is inconclusive (*26*). The small halos, spiraling into the Milky Way and other systems, should puff up the thin disks of normal galaxies to a greater degree than is observed (*27*, *28*)

The density profile of dark matter halos should exhibit a cuspy core in which the density rises sharply as the distance from the center decreases, in contrast to the central regions of many observed self-gravitating systems. Clusters of galaxies, as observed in studies of gravitational lensing, have less cuspy cores than do computed models of massive dark matter halos (*29*). Ordinary spiral galaxies have much less dark matter in their inner parts than expected (*30*, *31*), as do some low-surface-brightness galactic systems (*32*). Dwarf galaxies, like our companion systems Sculptor and Draco, have nearly uniform-density cores in contrast to the expected cuspy density profile (*33*, *34*). Hydrodynamic simulations produce galaxy disks that are too small and have too little angular momentum as compared to observations (*35*). Many high-surface-brightness spiral galaxies exhibit rotating bars, which are normally stable only if the core density is lower than predicted (*36*).

It is conceivable that the resolution of the growing list of problems lies in complex but more ordinary astrophysical processes. Numerous ingenious but conventional explanations for the absence of substructure have been proposed (*37–39*). The second set of objections, based on the cuspy density profile expected for CCDM, is observationally stronger, but here it may be that the theoretical predictions of a cuspy profile are not as certain as had been supposed (*40–42*). Overall, however, the evidence to date, taken in its totality, does indicate a discrepancy between the predicted high densities and the observed much lower densities in the inner parts of dark matter halos, ranging from those in giant clusters of galaxies [mass (M) $\geq 10^{15}$ solar masses ($M_\odot$)] to those in the smallest dwarf systems observed ($M \leq 10^9\ M_\odot$).

Alternatives to Cold Dark Matter

The possible discrepancies between theory and observation have motivated new proposals about the nature of dark matter. Each proposed variation from CCDM has two properties: (i) it can solve some or all of the problems described in the previous section, and (ii) it leads to additional predictions that would distinguish it from all the other alternatives. We discuss the following possible alternative models of dark matter.

1) Strongly self-interacting dark matter (SIDM). The dark matter might have a significant self-scattering cross-section σ, comparable to the nucleon-nucleon scattering cross-section (*43*). Then in any halo, large or small, where the number of particles per unit area (the surface density) $\times\ \sigma$ is greater than unity, collisions among the dark matter particles lead to a complex evolution of the structure. During the initial phase of this process, which lasts longer than the present age of the universe, the central densities decline in the desired fashion because of the scattering of dark matter particles. Also, scattering strips the halos from small clumps of dark matter orbiting larger structures, making them vulnerable to tidal stripping and reducing their number.

2) Warm dark matter (WDM). Dark matter may be born with a small velocity dispersion (for example, through decay of another species) (*44*, *45*), which leaves it with a velocity of perhaps only 100 m/s. Extrapolating back in time, this velocity increases to a value sufficient to have a significant effect on small-scale structure, because the particles are moving too fast to cluster on these scales. There are fewer low-mass halos, and all halos have a less steep profile in the innermost core. Also, because most of the lowest-mass halos are born from the fragmentation of larger structures in this picture, they are found in high-density regions, and the voids tend to be emptier of small systems than in the CCDM scenario.

3) Repulsive dark matter (RDM). Dark matter may consist of a condensate of massive bosons with a short-range repulsive potential (*46*). The inner parts of dark matter halos would behave like a superfluid and be less cuspy.

4) Fuzzy dark matter (FDM). Dark matter could take the form of ultralight scalar particles whose Compton wavelength (effective size) is the size of a galaxy core (*47*). Therefore, the dark matter cannot be concentrated on smaller scales, resulting in softer cores and smaller-scale structure.

5) Self-annihilating dark matter (SADM).

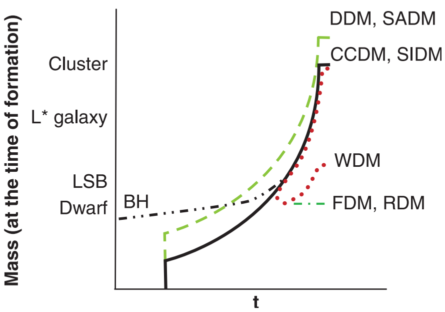


Fig. 2. History of structure formation: the time of formation for objects of a given mass (as measured at formation) for structures with increasing mass [dwarf galaxies, low-surface-brightness (LSB) galaxies, ordinary (L*) galaxies, and galaxy clusters] for different models of dark matter. Structure formation begins shortly after the onset of the matter-dominated epoch (left side).

Dark matter particles in dense regions may collide and annihilate, liberating radiation (*48*). This reduces the density in the central regions of clusters by direct removal of particles from the center and by the reexpansion of the remainder as the cluster adjusts to the reduced central gravity.

6) Decaying dark matter (DDM). If early dense halos decay into relativistic particles and lower mass remnants, then core densities, which form early, are reduced without altering large-scale structure (*49*).

7) Massive black holes (BH). If the bulk of the dark matter in galactic halos were in the form of massive black holes with masses of about one million $M_\odot$, then several dynamical mysteries concerning the properties of our galaxy could be better understood (*50*). In normal galaxies, dynamical friction between the massive black holes and the ordinary matter would cause the black holes in the central few kiloparsecs to spiral into the center, depleting those regions of dark matter and providing the ubiquitous central massive black holes seen in normal galaxies.

Determining the Nature of Dark Matter

At first sight, the conceivable alternatives to CCDM are so numerous that it may seem impossible ever to distinguish among them. However, each alternative produces distinctive modifications on small scales that can be tested through improved astronomical observations and numerical simulations. The local universe—the small objects that orbit galaxies and the galaxy cores—turns out to be a marvelous laboratory for examining the nature of dark matter.

SIDM, BH, or SADM only affect halos when the interaction rate rises above a certain threshold value. The interaction rate depends on the surface density if the cross-section is velocity-independent or, more generally, is the product of the cross-section and velocity. In all these cases, the interaction effect is slow because only a few scatterings take place within the lifetime of the universe. WDM, RDM, or FDM have a built-in characteristic length scale below which dark matter halos are affected. DDM has a characteristic built-in time scale after which dark matter halos are affected on all length scales and for all surface densities.

The alternatives also alter the history of structure formation compared to CCDM in different ways. SIDM maintains the same sequence of structure formation but slowly rearranges the distribution of dark matter in dense regions. SADM is similar, except that it removes dark matter altogether from dense regions. Depending on details, RDM and FDM may or may not affect the sequence of structure formation either, but they ensure that the smaller-scale objects are forced to have a low density. DDM removes dark matter on all scales beginning after a characteristic decay time; because a lot of mass is lost through the decays, a higher rate of clustering is required throughout to match the observed galaxy cluster masses and match the other proposals. WDM delays the onset of structure formation until the dark matter cools sufficiently to gravitationally cluster, initially suppressing small-scale structure formation but then creating it later by the fragmentation of larger-scale structures. Finally, the BH alternative requires that significant nonlinear structure on one million $M_\odot$ scales be built in ab initio, rather than grown from small fluctuations.

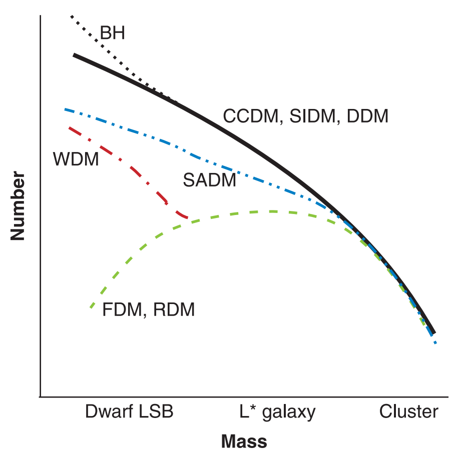


Fig. 3. Demography: how the number of objects of a given type depends on their mass (as observed today) for different dark matter models.

Because of these differences, the candidates for dark matter each face distinctive constraints and challenges. If the cross-section is too large, self-interaction or self-annihilation could lead to the evaporation of the halos of galaxies in clusters, which is in conflict with observations (*31*, *51*). For WDM, for which structure formation is delayed as compared to the standard picture, evidence for early galaxy and star formation provides a strong constraint. If the high electron-scattering optical depth found by WMAP is confirmed (an indicator of substantial star formation at very early epochs), there would not be room for any delay (*21*, *52*). Similarly, SADM could destroy all small halos made at early times before they become sites for new small galaxies. A challenge for DDM is that it requires a higher production of massive dense clusters in the early universe than has been observed in order to obtain the right mass distribution after decay.

We suggest that new kinds of observations may be able to distinguish among the candidates for dark matter by taking advantage of their qualitative differences. To be quantitative in our predictions, detailed numerical simulations of each case are necessary. It may be that some of the guesses we are putting forward will turn out to be incorrect when accurate calculations are made.

First we consider the epoch at which objects of different mass will form in the different scenarios (Fig. 2). To give the same structures today, objects of a given mass will need to form earlier in the DDM, SADM, and BH scenarios as compared to the standard CCDM and SIDM scenarios. The low-mass objects will form later in at least some FDM and RDM scenarios, and in the WDM scenario, they will form later and only from the fragmentation of more massive objects. The mass of, and even the existence of, low-mass galaxies at early times will provide a valuable diagnostic to distinguish among the alternatives.

Next we look at the demography: that is, how many small and large dark matter halos are expected in the local universe when population studies are completed (Fig. 3). In the

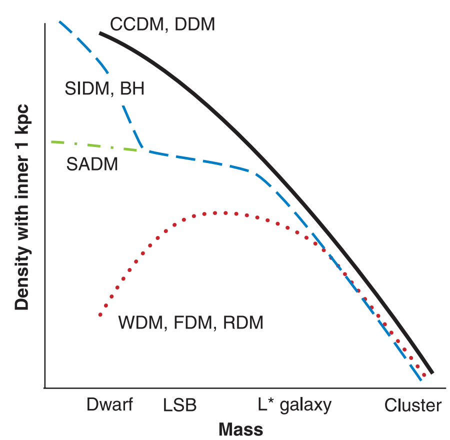


Fig. 4. Internal structure: how the density of the inner 1 kpc depends on the mass of the system for different dark matter models.

WDM, FDM, and RDM scenarios, low-mass objects are underabundant as compared to the CCDM, SIDM,and SADM scenarios; and in the BH scenario, they are probably overabundant. WDM calculations (*45*) reveal that objects made by fragmentation are present but at a lower level. The small halos may be difficult to observe directly, because they may be unable to retain gas long enough to make observable galaxies. But these small dark halos may be detected through their gravitational effects, such as lensing, puffing up of disks, and other dynamical interactions.

The internal structure of the halos provides another feature to distinguish one model from another. In the CCDM model, low-mass halos were made early when the universe was denser, and so they are more dense than structures formed later. This is shown in their internal structure. So, Fig. 4 reflects the historical conditions shown in Fig. 2 but allows one to study nearby objects. This is a critical issue because the inner parts of dark matter halos do seem to be considerably less dense than expected in the CCDM model. Here the BH scenario is complex. For isolated dark matter halos, which do not contain ordinary matter, the dynamical evolution will be qualitatively similar to that of star clusters. On a time scale proportional to the dynamical (or orbital) time multiplied by the ratio of the system mass to the typical black hole mass, the inner profile will first flatten and then collapse via a process called the gravothermal instability. For parameters appropriate to galactic dark matter halos, even the first process will only occur for the lowest-mass dwarf systems, and thus less cuspy cores would be expected in the local dwarf galaxies. In normal galaxies, the stronger interaction is between the black holes and the normal stellar component, and this leads, as noted before, to clearing out the black holes from the inner parts of the galaxies, with them sinking to the center where they either merge or are ejected.

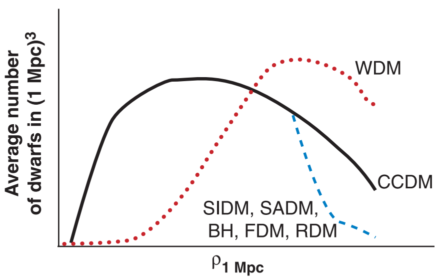


Fig. 5. Environment: how the number of dwarfs in a (1 Mpc)^3 volume depends on the average density within that volume.

Finally, in Fig. 5, we examine the environments within which different kinds of objects should be found. In the CCDM model, low-mass halos will be distributed relatively more uniformly than will the higher-mass halos, so that the large voids seen in the distribution of massive galaxies should be populated with halos of low mass and perhaps also with associated low-mass galaxies. To date, studies have not found such galaxies, but we do not yet know if this because of an absence of the predicted low-mass halos in the voids or simply because the ones that are there have not been able to make galaxies. In the WDM scenario, the low-mass halos are typically near the high-mass ones, because they form from the fragmentation of larger structures. For the SIDM, SADM, FDM, and RDM scenarios, the abundance of low-mass objects will decline in the vicinity of the highest-mass ones. In SIDM, it will be because interactions will boil away the cooler low-mass halos by direct particle-particle collisions, and in the other three cases, it is because the low-mass halos will have a low internal density and be fragile, hence easily shredded in tidal encounters with their bigger brothers. For the BH scenario, the voids would be heavily populated with small dark matter systems, but these might or might not contain observable stellar systems.

Conclusions

There are a variety of clues telling us that the universe may not be as simple as the CCDM model. Although the CCDM model is able to correctly predict observations made on the largest cosmological scales down to roughly those of galactic scale, and from the early universe to the present epoch, there are many indications that on subgalactic scales it predicts that there should be more dark matter than is detected gravitationally. Numerical simulations predict that all galaxies should contain cuspy cores, where the density of dark matter rises sharply with decreasing radius, and most observations do not confirm this prediction. We need more accurate simulations and more accurate observations to see whether these discrepancies are real. If they are, then there are several interesting suggestions that could account for the less cuspy cores and, more important, would lead to predictions of other observables that could be used to test the variant types of dark matter. These include the history of dark halo formation, the demography (mass distribution) of low-mass halos, the detailed interior density distribution of galaxy halos, and the environments within which different kinds of astronomical objects are found. We have sketched out the kinds of astronomical tests that could be done to narrow the search, but if history teaches us anything it is that the next important clues will come from a surprising direction. Some observation or calculation will be made that will reorient our inquiries and, if this happens as has happened so often in the past, we will realize that the important evidence has been sitting unnoticed under our noses for decades.

References and Notes

1. F. Zwicky, *Astrophys. J.* **86**, 217 (1937).
2. M. S. Roberts, A. H. Rots, *Astron. Astrophys.* **26**, 483 (1973).
3. J. P. Ostriker, P. J. E. Peebles, A. Yahil, *Astrophys. J. Lett.* **193**, L1 (1974).
4. J. Einasto, A. Kaasik, E. Saar, *Nature* **250,** 309 (1974).
5. V. C. Rubin, N. Thonnard, W. K. Ford Jr., *Astrophys. J. Lett.* **225**, L107 (1978).
6. A. H. Guth, *Phys. Rev. D* **23**, 347, (1981).
7. R. P. Kirshner, *Science* **300**, 1914 (2003).
8. P. J. E. Peebles, *Astrophys. J.***284,** 439 (1984).
9. G. Efstathiou, W. J. Sutherland, S. J. Maddox, *Nature* **348**, 705 (1990).
10. L. Krauss, M. S. Turner, *Gen. Relativ. Gravit.* **27**, 1137 (1995).
11. J. P. Ostriker, P. Steinhardt, *Nature* **377**, 600 (1995).
12. A. Reiss *et al.*, *Astron. J.* **116**, 109 (1998).
13. S. Perlmutter *et al.*, *Astrophys. J.* **517**, 565 (1999).
14. C. L. Bennett *et al.*, preprint available at http://lanl.arXiv.org/abs/astro-ph/?0302207 (2003).
15. L. Wang, R. R.Caldwell, J. P. Ostriker, P. J. Steinhardt, *Astrophys. J.* **530**, 17 (2000).
16. J. Bardeen, P. J. Steinhardt, M. S. Turner, *Phys. Rev. D* **28**, 679 (1983).
17. A. H. Guth, S.-Y. Pi, *Phys. Rev. Lett.* **49**, 1110 (1982)
18. S. W. Hawking, *Phys. Lett. B* **115**, 295 (1982).
19. V. F. Mukhanov, G. V. Chibisov, *J. Exp. Theor. Phys. Lett.* **33**, 532 (1981).
20. A. A. Starobinskii, *Phys. Lett. B* **117**, 175 (1982)
21. D. N. Spergel *et al.*, preprint available at http://lanl.arXiv.org/abs/astro-ph/?0302209 (2003).
22. M. C. Begelman, *Science* **300**, 1898 (2003).
23. J. F. Navarro, C. S. Frenk, S. D. M. White, *Astrophys. J.* **490,** 493 (1997).
24. B. Moore, F. Governato, T. Quinn, J. Stadel, G. Lake, *Astrophys. J. Lett.* **499,** L5 (1998).
25. A. V. Kravtsov, A. A. Klypin, J. S. Bullock, *Astrophys. J.* **502,** 48 (1990).
26. N. Dalal, C. S. Kochanek, *Astrophys. J.* **572**, 25 (2002).
27. G. Toth, J. P. Ostriker, *Astrophys. J.* **389**, 5 (1992).
28. A. S. Font, J. F. Navarro, J. Stadel, T. Quinn, *Astrophys. J.* **563**, L1 (2001).
29. J. A. Tyson, G. P. Kochanski, I. P. Dell'Antonio, *Astrophys. J. Lett.* **498**, L107 (1998).
30. J. J. Binney, N. W. Evans, *Mon. Not. R. Astron. Soc.* **327**, L27 (2001).
31. R. Davé, D. N. Spergel, P. Steinhardt, B. Wandelt, *Astrophys. J.* **547**, 574 (2001).
32. F. C. Van den Bosch, E. B. Bobertson, J. J. Dalcanton, W. J. G. de Bok, *Astron. J.* **119**, 1579 (2000).
33. F. Stoehr, S. D. M. White, G. Tormen, V. Springel, *Mon. Not. R. Astron. Soc.* **335**, L84 (2002).
34. J. T. Kleyna, M. Wilkinson, G. Gilmore, W. N. Evans, preprint available at http://lanl.arXiv.org/abs/astro-ph/?0304093 (2003).
35. J. F. Navarro, M. Steinmetz, *Astrophys. J.* **528**, 607 (2000).
36. V. P. DeBattista, J. A. Sellwood, *Astrophys. J.* **493**, L5 (1998).
37. J. Silk, *Astrophys. J.* **211**, 638 (1977).
38. W. A. Chiu, N. Y. Gnedin, J. P. Ostriker, *Astrophys. J.* **563**, L21 (2001).
39. K. Nagamine, M. Fukugita, R. Cen, J. P. Ostriker, *Mon. Not. R. Astron. Soc.* **327**, L10 (2001).
40. C. Power *et al.*, *Mon. Not. R. Astron. Soc.* **338**, 14 (2003).
41. S. Ghigna *et al.*, *Astrophys. J.* **544**, 616 (2000).
42. M. Ricotti, preprint available at http://lanl.arXiv.org/abs/astro-ph/?0212146 (2002).
43. D. N. Spergel, P. J. Steinhardt, *Phys. Rev. Lett.* **84**, 3760 (2000).
44. P. Colín, V. Avila-Reese, O. Valenzuela, *Astrophys. J.* **542**, 622 (2000).
45. P. Bode, J. P. Ostriker, N. Turok, *Astrophys. J.* **556**, 93 (2001).
46. J. Goodman, *New Astronomy* ***5*****, 103 (2000).**
47. W. Hu, R. Barkana, A. Gruzinov, *Phys. Rev. Lett.* **85**, 1158 (2000).
48. L. Kaplinghat, L. Knox, M. S. Turner, *Phys. Rev. Lett.* **85**, 3335 (2000).
49. R. Cen, *Astrophys. J.* **546**, L77 (2001).
50. C. Lacey, J. P. Ostriker, *Astrophys. J.* **229**, 633 (1985).
51. J. F. Hennawi, J. P. Ostriker, *Astrophys. J.* **572**, 41 (2002).
52. N. Yoshida, V. Springle, S. D. M. White, G. Tormen, *Astrophys. J. Lett.* **544**, L87-L90 (2000).

Throwing Light on Dark Energy

Robert P. Kirshner

Supernova observations show that the expansion of the universe has been speeding up. This unexpected acceleration is ascribed to a dark energy that pervades space. Supernova data, combined with other observations, indicate that the universe is about 14 billion years old and is composed of about 30% matter and 70% dark energy. New observational programs can trace the history of cosmic expansion more precisely and over a larger span of time than has been done to date to learn whether the dark energy is a modern version of Einstein's cosmological constant or another form of dark energy that changes with time. Either conclusion is an enigma that points to gaps in our fundamental understanding of gravity.

Observations of exploding stars halfway back to the Big Bang reveal a surprising phenomenon: The expansion of the universe has been speeding up in the past 7 billion years. We attribute this effect to the presence of a dark energy, whose energy density helps make the universe flat and whose negative pressure produces cosmic acceleration. On the basis of observations of supernova brightness, of the dark matter that makes galaxies cluster, and of the angular scale of primordial freckles in the glow from the cosmic microwave background (CMB), we infer that about 28% of the universe is matter and 72% is dark energy. In the self-proclaimed age of "precision cosmology," we know the amount of each component to a few percent, but in the spirit of "honest cosmology" we also have to admit we do not know precisely what either of them is. But we are not helpless. We can observe light emitted by supernova explosions to trace the history of cosmic expansion to learn more about the invisible forces that shape the universe.

Evidence for the nature of the dark energy comes from the observed brightness of a particular class of supernova explosions called type Ia supernovae (SN Ia's). Defined empirically from their spectra (*1*), these events mark the thermonuclear destruction of white dwarf stars. A white dwarf, stable when solitary up to 1.4 solar masses, can accrete matter from a companion when it is in a binary system. A white dwarf in a binary will explode violently, destroying the star, when accreted mass provokes the carbon and oxygen in its interior to erupt in a runaway thermonuclear explosion (*2*, *3*). SN Ia's are infrequent events, erupting roughly once per century in a galaxy, and found in all types of galaxies. SN Ia's are useful for probing the history of cosmic expansion and the nature of dark energy because they are very bright, typically about 4×10^9 times the luminosity of the Sun. With careful measurements of the color and the apparent brightness during the month when a SN Ia shines most brightly, the distance to an individual explosion can be derived to better than 10% (*4–6*). This precision makes SN Ia's the best standard candles in extragalactic astronomy: Observations of nearby and bright SN Ia's help determine the present rate of cosmic expansion, the Hubble constant, H_o, of 72 ± 8 km s^{-1} Mpc^{-1} (*7*, *8*). Observations of the brightness and spectra of these objects measure the relation between distance and redshift for the universe. The redshifts of supernovae at different distances reveal changes in the rate of cosmic expansion that have developed while the light was in flight to us from explosions over 7 billion light-years away. The observed effect is that supernovae at a redshift of $z = 0.5$ (roughly one-third of the way back to the Big Bang) appear about 25% dimmer than they would in a universe without cosmic acceleration: Acceleration increases the distance the light must travel to reach us.

The first clues from distant supernovae were contradictory (*9–11*), but, by 1998, evidence from supernova distances favored a universe that was accelerating (*12*, *13*). Present work includes a widening stream of supernova discoveries at low redshift (*14*), diligent follow-up (*15*), and a growing body of well-observed cases to compare with the high-redshift data (*16*, *17*). In addition, recent results (*18*) independently confirm the 1998 results, whereas the analysis of supernovae and their host galaxies (*19*) showed persuasively that uncorrected extinction by dust in galaxies, a possible source of systematic error, most likely does not produce the observed dimming of distant SN Ia's.

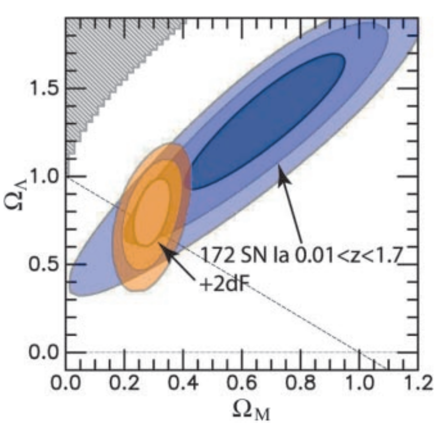


Fig. 1. The concordance diagram. In blue, 1-, 2-, and 3σ confidence contours for Ω_Λ and Ω_m based on 172 SN Ia's from (*18*). The smaller orange error ellipses result from combining supernova data with information from large-scale structures in a flat universe (*49*, *50*).

The published sample of high-z supernovae has now been extended to the decisive redshift range of $z \sim 1$ (*18*, *20*, *21*), where the effects of cosmology begin to change sign from making supernovae dim to making them a little brighter than they would otherwise appear. These observations sample directly the epoch when the balance between dark energy and dark matter tilted from cosmic deceleration because of dark matter to cosmic acceleration caused by dark energy. This opens the prospect of learning how the dark energy behaves as the universe expands on the basis of careful observations in the era at the onset of cosmic acceleration.

Improved evidence for dark energy from supernovae has boosted these results from a startling possibility to conventional wisdom in just the past 5 years. The general acceptance of this new picture of a universe dominated by dark energy derives from the neat fit of supernova data with other cosmological measurements, including galaxy clustering as a measure of dark matter, the ages of stars, and measurements of the CMB. Each of these strands in the web of inference has grown more secure, and the pleasant result has been a trend toward greater concordance from independent directions. These results converge on a universe that is 13.6 ± 1.5 billion years old and expanding at a present rate of 72 ± 8 km s^{-1} Mpc^{-1}, which is composed of 28 ± 5% matter and 72% dark energy (*18*, *22*) (Fig. 1).

Nature of the Dark Energy

One possible explanation is that dark energy is the modern version of Einstein's cosmological constant (*23–25*). In 1917, Einstein introduced a curvature term to produce static, eternal solutions to his field equations, in accord with the view then current that the Milky Way was the entire universe and the observational fact that the motions of its stars showed no systematic expansion or contraction. Legend holds that Einstein, after learning of Hubble's work on cosmic expansion based on galaxies outside the Milky Way, smote himself on the forehead and declared the cosmological constant his greatest

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 01238, USA. E-mail: kirshner@cfa.harvard.edu

blunder. This phrase does not occur in any of Einstein's writings but is derived from a line in Gamow's autobiography, in which Gamow, describing his own early studies of general relativity in St. Petersburg, says that "much later" Einstein called the cosmological constant "perhaps the biggest blunder of my life" (*26*). Einstein's own comments, written with the astronomer DeSitter, are more sensible than Gamow's legend. In 1932, they wrote about the cosmological constant: "An increase in the precision of data derived from observations will enable us in the future to fix its sign and determine its value" (*27*). But there isn't any doubt that Einstein felt the cosmological constant was repugnant as well as repulsive. In a 1947 letter to Lemaitre, he wrote, "Since I introduced this term, I had always a bad conscience. . .. I am unable to believe that such an ugly thing should be realized in nature" (*28*).

Following Zel'dovich (*29*, *30*), the modern interpretation of the cosmological constant regards it not as a curvature but as a vacuum energy density (*31*). This vacuum energy has quite unintuitive properties, most notably a negative pressure, P. If the vacuum energy density is really constant, then if you imagine a cylinder bounded by a piston with this stuff in it, and you wish to expand the volume by an increment dV, you will need to pull on the piston to do an amount of work PdV that will result in an increased energy inside the cylinder (because the energy density stays constant) (*32*). This negative pressure has important consequences for cosmic expansion, expressed in the standard Friedmann equations for the cosmic scale factor, $a(t)$, which describes the evolution of distances between galaxies in the universe (*33*). The gravitational acceleration in general relativity, which determines the sign of the second time derivative of a, a'', depends on the quantity $\rho c^2 + 3P$, where c is the speed of light. Matter has positive pressure (and very little of it in the present universe), which, along with positive density, ensures that a universe made of matter will always decelerate. But a cosmological constant can produce a negative pressure that changes the sign of $\rho c^2 + 3P$ to produce repulsive effects as long as $P < -1/3\rho c^2$.

In 1917, Einstein chose a value for the pressure that made the universe static, but this was an unstable equilibrium. For the cosmological constant (or any dark energy that changes slowly enough as the universe expands), P is negative and effectively constant. This makes an expanding universe accelerate: As the matter density decreases, the negative pressure does not, and eventually this will make the universe expand exponentially. In 1932, Arthur Eddington did not think the cosmological constant was a blunder; he thought the observed Hubble expansion might well be just the first-order view of a universe accelerating from rest because of a cosmological constant (*34*). The 1998 supernova results point to a dark energy that has negative pressure, so that galaxies separating after the Big Bang and gently decelerated by dark matter for the next 7 billion years are presently accelerating exponentially away from one another.

Although there is no particular conceptual problem with dark energy in the form of either a cosmological constant or some other energy that changes slowly with time, there are two serious quantitative problems. The data require a dark energy, which can be expressed as a fraction of the energy density of the universe as $\Omega_\Lambda = 0.7$ (*35*). One theoretical problem this poses is that the natural scale for the energy of the vacuum for gravitation is set by the Planck mass (M_{Planck}) at $\rho_{\mathrm{vacuum}} = M_{\mathrm{Planck}}{}^4\, c^3\, h^{-3}$ (where h is the Planck constant) which is 120 orders of magnitude larger than the astronomically observed value. This discrepancy can be ameliorated by cutting off the energy scale at the point where current knowledge of high-energy physics fades, but we are still left with a 55-orders-of-magnitude difference between theory and observation (*36*).

Another quantitative theoretical problem is that the present value of Ω_Λ implies that 70% of the energy in the universe is now in the form of dark energy. The sum of Ω_Λ and Ω_{matter} stays the same as the universe expands: If it is 1.000 today, it was 1.000 yesterday and will be 1.000 tomorrow. But the ratio $\Omega_\Lambda/\Omega_{\mathrm{matter}}$, about 2 today, changes briskly as the universe expands, because the vacuum energy stays constant whereas the mass density scales as a^{-3}. Even a modest exploration of the past, back to redshift $z \sim 1$, where $a^3 \sim (1 + z)^3$ is 8, means we will be looking back to the regime where dark matter dominated the balance of cosmic energy by as much as dark energy does today. The shift about 7 billion years ago from a decelerating universe dominated by dark matter to an accelerating universe dominated by dark energy means we just happen to live at the unique moment when this is true. When the universe attains twice its current age, we will have $\Omega_\Lambda/\Omega_{\mathrm{matter}} \sim 10$, and, when it was half its current age, we had $\Omega_\Lambda/\Omega_{\mathrm{matter}} \sim 1/10$. Why do we live at exactly the moment (where "moment" means a span from 7 billion years in the past to 14 billion years in the future) when the vacuum energy is about the same as the mass energy density? Nobody knows. Einstein thought the cosmological constant was ugly, and, in their hearts, modern theoretical physicists agree, but the astronomical observations seem persuasive that the universe is constructed in this extravagant way and that this problem cannot be wished away. Other forms of dark energy that change over time in a different way can avoid this problem and have been proposed (*37–40*).

Observing the Era of Acceleration

Although theorists are bothered by the coincidence of our era with the shift from a decelerating universe to an accelerating one, observers are delighted. Because this change is recent, it is potentially within view, and, by using the best of current technology, it provides a direct test of whether unforeseen systematic shifts in the intrinsic luminosity of supernovae are producing an illusion of cosmic acceleration. If unaccounted-for dust, or changes in the ages of stars, or drifts in the chemical composition of stars, rather than cosmology, make distant supernovae dim, then going to higher redshift should exacerbate those problems and make them fainter still. But, if the universe shifted from deceleration to acceleration at some time in the not-too-distant past, we would expect the sign of the effect on apparent magnitude to change. SN Ia's at $z \sim 0.5$ are dimmed by the effect of cosmic expansion, but we should expect SN Ia's beyond redshift 1 to appear a little brighter than they would otherwise if the universe were decelerating at the epoch of their detonation. This is a test that the supernova observations could fail.

The observational problems of finding and measuring supernovae at $z \sim 1$ are challenging. Because the entire spectrum is redshifted by a factor of $1 + z$, this means that the ordinary visible wavelength bands of optical astronomy provide measurements of the ultraviolet light emitted by SN Ia's, whereas the bulk of the flux is received at longer wavelengths. Large arrays of silicon-based charge-coupled devices (CCDs), such as the MOSAIC camera at Cerro Tololo Inter-American Observatory, the SUPRIME camera at Subaru, or the MEGACAM at the Canada-France-Hawaii Telescope (CFHT), are today's best tools for supernova searches. By searching in the reddest bands where these devices work well, in the range from 800 to 900 nm, and increasing the exposure times enough to detect objects with apparent magnitudes in the I band ~ 24 magnitude, a search can be tuned to emphasize the high-z supernovae, as reported by (*18*). Obtaining spectra of these most distant objects to get the redshift and to confirm that the object is a SN Ia is also a challenge. Because the brightness of the supernova is only a few percent of the brightness of the night sky, it typically takes hours of integration with the largest telescopes, such as Keck, Gemini, or the European Southern Observatory's Very Large Telescope (VLT), to obtain spectra of these faint objects. Photometry from the ground requires precise subtraction of the background galaxy, which is typically several times brighter than the SN Ia. This can be done from the ground, but Hubble Space Telescope (HST) observations, with their exquisite resolution, are much easier to use. The evidence to date (Fig. 2), though slim at $z \sim 1$, favors the view that we are seeing past the era of acceleration at $z \sim 0.5$, back to the timeof cosmic deceleration near $z \sim 1$.

Searches from the ground have the advantages of large telescope apertures (Subaru, for example, has 10 times the collecting area of HST) and large CCD arrays [the CFHT has a 378-million-pixel camera, compared to the new Advanced Camera for Surveys (ACS) on HST, which has 16 million pixels]. The advantages of space include avoiding the bright and variable night sky encountered in the near-infrared; the potential of much sharper imaging for point sources, like supernovae, to distinguish them from the galaxies in which they reside; and better control over the observing conditions, which need not factor in weather and moonlight. During December 1997 and into early 1998, a repeat exposure of the Hubble Deep Field (HDF) was carried out (*41*, *42*), followed by repeated imaging with its infrared camera. Without knowing it, the infrared camera team had selected as their target field the site of SN 1997ff, which was subsequently recognized and extracted from the data archive (*43*). The observations do not include a spectrum of either the supernova or the galaxy, but the observed colors were used to estimate the redshift at $z = 1.7 \pm 0.1$. The apparent magnitude of SN 1997ff is about 1 full magnitude brighter than expected in a universe with no acceleration or deceleration. Although nobody regards SN 1997ff as a conclusive demonstration of cosmic deceleration, the data are in good accord with what would be expected if the universe really did change from deceleration to acceleration. If many such objects could be measured well, and they traced the expected path in the plot of apparent magnitude and redshift, they would tell us whether we are really seeing back to the age of cosmic deceleration (*44*).

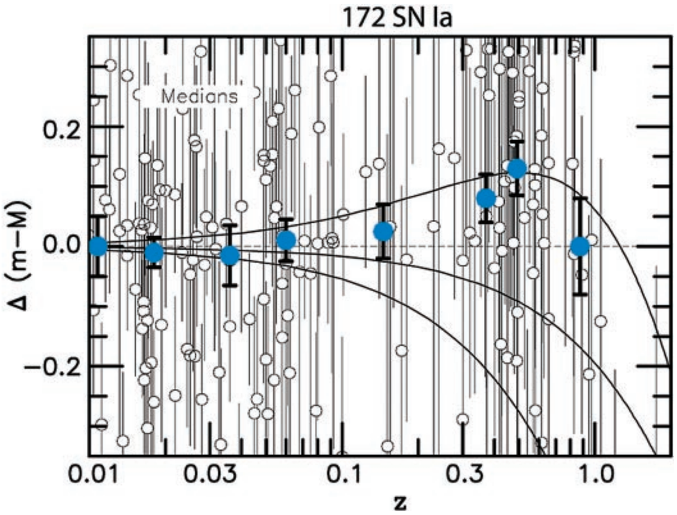


Fig. 2. Residual Hubble diagram: apparent magnitude difference between the expected magnitude in an empty universe and the observed magnitude of supernovae at each redshift (*18*). Individual points are shown with their quoted error bars. For clarity, medians in redshift bins are shown in blue. The theoretical lines shown correspond to Ω_Λ, Ω_m pairs of (0.7, 0.3), (0, 0.3), and (0, 1). The highest redshift bin may show signs of cosmic deceleration, as predicted by the top line.

The installation of ACS on HST has made it practical to search for supernovae with HST itself. The new camera has twice the area on the sky, sampling of the images that is twice as fine, and throughput that is five times better than HST's previous imager. In an early test, two SN Ia's at $z = 0.47$ (SN 2002dc) and $z = 0.95$ (SN 2002dd) were discovered with HST, which subsequently gathered beautiful light curves and spectra (*21*). The Great Observatories Origins Deep Survey (GOODS) program to reimage the HDF with ACS was optimized to detect transient events, especially high-redshift supernovae, by adopting a different approach to scheduling. Instead of relentlessly observing the HDF for 342 images over 10 days, as done in 1995, successive GOODS observations were spaced by 45 days, providing 5 epochs of data on two fields, HDF north and south. Whereas the GOODS team adds these images to build a superdeep field, the Higher-*Z* Team, led by Adam Riess (but with an active cast of dozens), subtracted the accumulated template image from each incoming frame. The Higher-*Z* Team has detected 42 supernovae, with redshifts ranging from $z = 0.3$ to $z = 1.8$, and 10 of these have $z \geq 1$ (*45*, *46*). When these exquisite data are fully analyzed, we can expect a much firmer report from the epoch of cosmic deceleration (Fig. 3). The HST is a powerful tool for discovery and measurement of supernovae that are too difficult to find and follow from the ground.

The Essence of Things

The era of cosmic acceleration is quite recent. This means that the observed effect of dimming SN Ia's has its largest amplitude in the relatively easily observed range from $z = 0.3$ to $z = 0.7$, where most of the present sample of high-z supernovae has already been accumulated. Tonry *et al.* (*19*) analyzed data for 230 SN Ia's with redshifts and distances. Most of these are in the nearby universe ($z < 0.1$), where the effects of acceleration are too subtle to detect: The signal to determine the best value of Ω_Λ comes from higher redshifts. The typical internal errors on the measurement of distance for each supernova are larger than we get for the best observed cases (Fig. 2). It would be good to construct a larger, more uniform sample with smaller errors.

If the dark energy is the cosmological constant, we know precisely what to expect for its behavior: The energy density remains unchanged. However, dyspepsia caused by the cosmological constant is strong enough to inquire whether the dark energy could have some other nature. For example, if the dark energy comes from some slow-changing energy field, as in quintessence models (*37–40*), then it would be of great interest to determine the properties of that field in the manner advocated by Einstein and DeSitter: from observation.

A simple parameterization of the possible forms of dark energy uses the idea of the cosmic equation of state (*47*). Suppose the dark energy density changes with the scale factor, a, as a power law $\rho \sim a^{-n} \sim (1 + z)^{-3(1 + w)}$. Here, w is the effective equation-of-state index, because by examining the way pressure changes with cosmic expansion, you can write an equation of state that connects the energy density to the pressure: $p = w\rho c^2$. Familiar values of w include $w = 0$, for ordinary matter and for

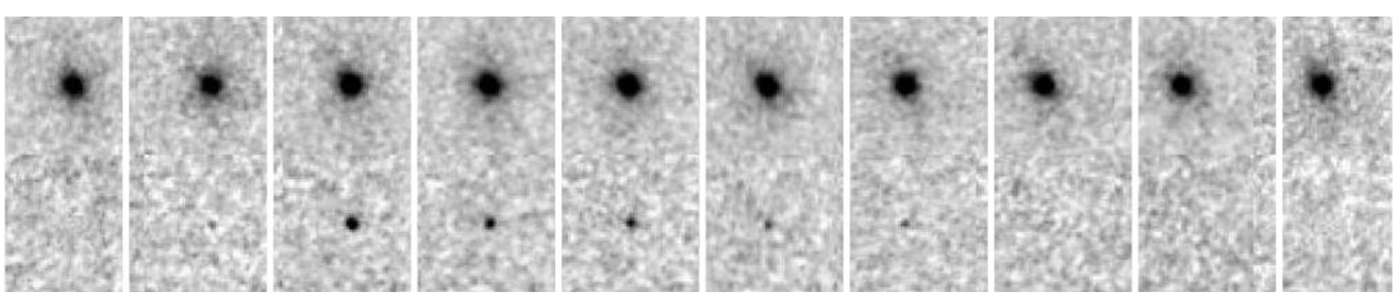


Fig. 3. The rise and fall of Thoth (SN 2002hp), a high-redshift supernova discovered and observed with the ACS on the HST by the Higher-*Z* Team.

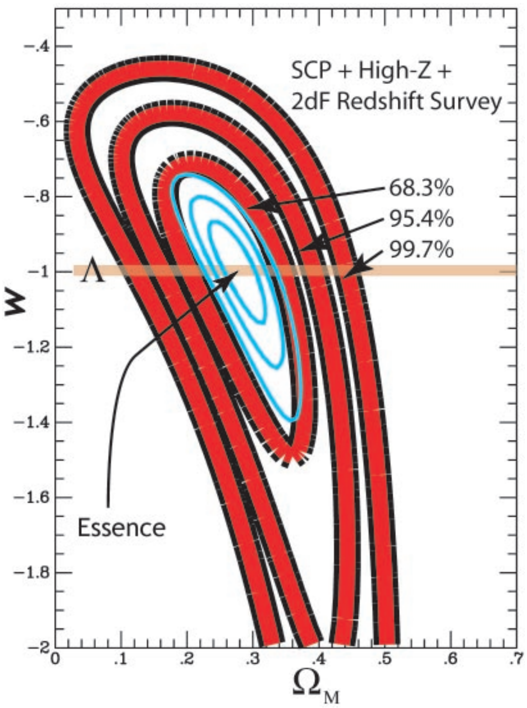


Fig. 4. The cosmic equation of state. Outer contours (red) give the current constraints from SN Ia's according to (*18*). Inner contours (blue) show the expected improvement in precision to be expected from completing the 200–SN Ia's catalog of the ESSENCE program.

cold dark matter that just gets diluted by expansion, and $w = 1/3$, for radiation that gets diluted and redshifted. For a true cosmological constant, $w = -1$. Other forms of dark energy might have different values of w that could be determined from careful observations of the onset of acceleration. On the basis of the 1998 supernova observations, the cosmic equation of state is consistent with $w = -1$ (*48*), but the precision of these early results was not very high. The current state of the art based on combining supernovae with constraints from galaxy redshift surveys (*19*, *49*, *50*) is shown in Fig. 4. The observed constraints in the $\Omega_m - w$ plane assume that $\Omega_m + \Omega_\Lambda = 1$. The data favor a value of $\Omega_m = 0.28$, consistent with independent methods (*49*, *50*) and also consistent with a value of $w = -1$. The 95% confidence interval on w is formally in the range $-1.48 < w < -0.72$. If we are bold enough to assert that $w > -1$, which seems sensible enough on the basis of energy conditions [but see (*51*) for an exploration of what it might mean to have $w < -1$], then the 95% confidence upper limit on w is $w < -0.73$. These constraints are similar to those reported using results from the Wilkinson Microwave Anisotropy Probe (WMAP) satellite, where the early results give $w < -0.78$ at 95% confidence (*22*).

So far, so good. But a larger, more homogeneous data set would have the potential to do much better at this investigation of the nature of the dark energy. A program to build that data set, dubbed ESSENCE [Equation of State: SupErNovae trace Cosmic Expansion (*52*); pronounced "SNs"], is under way at the Cerro Tololo Inter-American Observatory. With the use of a powerful data pipeline developed by Chris Stubbs of the University of Washington and a wide range of collaborators from the High-*Z* Team, led by Chris Smith and Nick Suntzeff at Cerro Tololo, the program aims to find and measure 200 SN Ia's in the redshift range of interest, $0.15 < z < 0.7$, in the next 5 years. Substantial spectroscopic backup to the program, to get the redshifts and to assure that the objects are really SN Ia's, comes from the use of Gemini, Magellan, VLT, Keck, and MMT Observatory. Figure 5 shows a sample of spectra obtained with the use of the Gemini spectrograph from the past year's observations. The inner contours of Fig. 4 show the expected improvement in the precision of measuring w that will result from completing the full ESSENCE program by 2006. This observing program cannot fail to be interesting. Either the contours will shrink around $w = -1$, in which case the cosmological constant will be an even stronger candidate for the dark energy, or they will converge on some other value that is different from –1, which would be even more exhilarating.

On the other hand, just learning the value of w is not the whole story on the dark energy. As several authors have pointed out (*53*, *54*), there are many conceivable forms of the dark energy, and no conceivable set of observations will rule out all the devious constructions of unchecked theoretical imagination.

What Next?

Supernovae have led the way in revealing cosmic acceleration. Quantitatively, the results agree with the independently measured values for Ω_m from large-scale structure and the result for $\Omega_m + \Omega_\Lambda$ from the CMB. The supernova results also place a strict limit on the cosmic age that fits with other lines of evidence. From the Tonry *et al.* compilation (*19*), if $w = -1$, then $H_o t_o$, the dimensionless expansion time, is 0.96 ± 0.04. For a value of $H_o = 72$ km s^{-1} Mpc^{-1} based on Cepheids and SN Ia's (*9*), this makes the elapsed time since the Big Bang, taking into account both the bygone era of deceleration and the modern era of acceleration, $13.6 \times 10^9 \pm 1.5 \times 10^9$ years. This is in good accord with an age of 12.5×10^9 years from 17 metal-poor globular clusters (*55*). If these systems began to form around $z = 8$, which corresponds to an incubation time of 0.6×10^9 years, this gives a cosmic age based on stellar evolution of 13.1×10^9 years. The expansion age from supernovae is also in good accord with the age inferred from WMAP of $13.7 \times 10^9 \pm 0.2 \times 10^9$ years (*23*)

All of this good news should not be a source of complacency. There are many aspects of SN Ia's that are poorly understood and that could affect their use as cosmic yardsticks in subtle ways (*56*). We do not know which stars become SN Ia's, and there may be a mix of supernovae of different types in any sample that we are lumping together and treating in the same way. We do not know how the chemical evolution of the parent population and the white dwarfs they form affects the luminosity of the supernovae they produce or how this should vary over time (*57*).

All of this is hidden beneath the surface and may create a floor of systematic variation that cannot be eliminated simply by increasing the sample size. One good path forward is to continue the discovery and study of SN Ia's in nearby galaxies, which includes a wide range of local chemical abundances and star-formation histories.

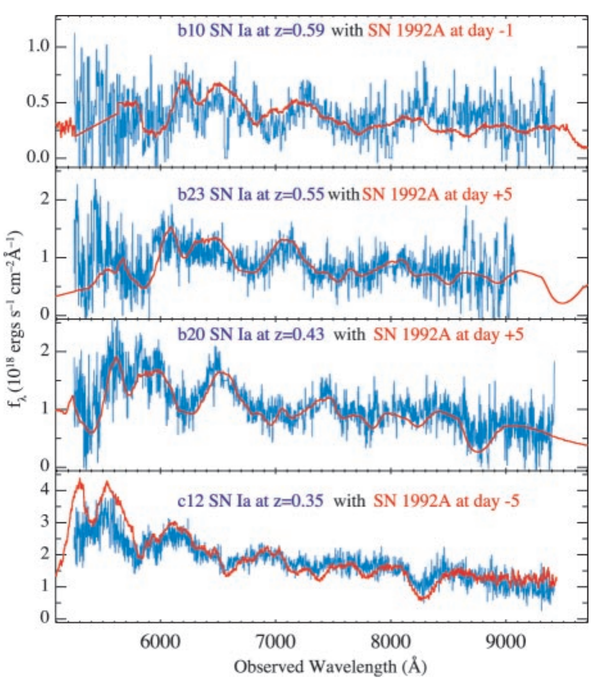


Fig. 5. Spectra of ESSENCE SN Ia's (blue) compared to a well-observed low-redshift SN Ia, SN 1992A (red). These data, from the Gemini Multi-Object Spectrograph at the Gemini south, show that the spectra of distant SN Ia's are well matched by nearby objects.

The present stream of discoveries from the Katzman Automatic Imaging Telescope (82 very nearby supernovae in 2002 alone), plus the valuable contributions of dedicated amateurs coupled with dogged follow-up, is the way forward. We know that the use of the light-curve shape helps decrease the scatter in supernova Hubble diagrams, and we may find that spectra help too. The really good photometric sample at low z now numbers $\sim$100 objects (*16*), and a spectroscopic sample of 845 spectra of 67 SN Ia's has been compiled (*58*). For the near term, we can use these data sets to investigate systematic effects. Larger samples will be forthcoming from the Legacy Survey (*59*) at CFHT and from the SN Factory (*60*) if they can provide adequate follow-up. These surveys will find fainter supernovae than the nearby searches. Follow-up will require a much larger investment of time to yield light curves and spectra of the same quality as those that can be observed for the nearby objects. The comparison of truly distant supernovae to the nearby sample shows no obvious differences in their spectra (*61*), as illustrated for some ESSENCE supernovae (Fig. 4), but pushing the systematic variations below 5% will require understanding subtle differences among the SN Ia's. All of this will have to come from semi-empirical work. First-principles computation of SN Ia explosions, luminosities, and spectra is, at present, too crude to predict directly the variations with epoch.

Rapid progress in measuring the CMB has come from a variety of approaches, including ground-based observations from high desert sites and from the South Pole, balloons, and WMAP. In the future, this field will be further advanced by elaborate satellites like Planck. In a similar way, the sustained observation of nearby supernovae, ground-based programs like ESSENCE, and straightforward extrapolation of the Higher-*Z* program on HST are certain to make progress in constraining dark energy. A wide-field imager, to make HST a truly formidable supernova harvester in an extended mission (*62*) and a quick, ruthlessly simple satellite could gain some of the needed data and sharpen the questions for the field in just a few years. The program described by the Supernova/Acceleration Probe (SNAP) collaboration (*63*) would be an extraordinary leap beyond these modest ideas. They propose a formidable 2-m telescope (about the size of HST) with a billion-pixel detector (120 times the size of ACS on HST) and an infrared spectrometer of unprecedented efficiency dedicated to supernova studies. The idea is to measure thousands of supernovae with excellent control of the systematics to reveal the fine details of cosmic acceleration and to infer more thoroughly the properties of the dark energy.

Theorists may be wary of the coincidence between the present and the onset of cosmic acceleration. Observers are delighted by this coincidence and by the coincidence between our own brief lives and the instant when technology has made these measurements possible. We are incredibly lucky to be working just at the moment when the pieces of the cosmic jigsaw puzzle are falling into place, locking together, and revealing the outline of the pieces yet to come. Dark energy is the biggest missing piece and a place where astronomical observations point to a gaping hole in present knowledge of fundamental physics.

References and Notes

1. A. V. Filippenko, *Ann. Rev. Astron. Astrophys.* **35**, 209 (1997).
2. J. Whelan, I. Iben, *Astrophys. J.* **186**, 1007 (1973).
3. K. Nomoto, F.-K. Thielemann, J. C. Wheeler, *Astrophys. J.* **297**, L23 (1984).
4. M. M. Phillips, *Astrophys. J.* **413**, L105 (1993).
5. A. G. Riess, W. H. Press, R. P. Kirshner, *Astrophys. J.* **473**, 88 (1995).
6. M. M. Phillips *et al. Astron. J.* **118**, 1766 (1999).
7. W. Freedman *et al.*, *Astrophys. J.* **553**, 47 (2001).
8. B. R. Parodi *et al.*, *Astrophys. J.* **540**, 634 (2000).
9. S. Perlmutter *et al.*, *Astrophys. J.* **483**, 565 (1997).
10. P. Garnavich *et al.*, *Astrophys. J.* **493**, 53 (1998).
11. S. Perlmutter *et al.*, *Nature* **391**, 51 (1998).
12. A. G. Riess *et al.*, *Astron. J.* **116**, 1009 (1998).
13. S. Perlmutter *et al.*, *Astrophys. J.* **517**, 565 (1999).
14. A. V. Filippenko *et al.*, in *Small Telescope Astronomy on Global Scales*, W. P. Chen, C. Lemme, B. Paczynski, Eds. (Astronomical Society of the Pacific, San Francisco, CA, 2001), pp. 121–130; available online at http://astron.berkeley.edu/~bait/kait.html.
15. S. Jha, thesis, Harvard University (2002).
16. M. Hamuy *et al.*, *Astron. J.* **112**, 2398 (1996).
17. A. G. Riess *et al.*, *Astron. J.* **116**, 1009 (1998).
18. J. Tonry *et al.*, *Astrophys. J.*, in press; preprint available online at http://arxiv.org/abs/astro-ph/0305008.
19. M. Sullivan *et al.*, *Mon. Not. R. Astron. Soc.* **340**, 1057 (2003); available online at http://arxiv.org/abs/astro-ph/0211444.
20. A. G. Riess *et al.*, *Astrophys. J.* **560**, 49 (2001).
21. J. P. Blakeslee *et al.*, *Astrophys. J.* **589**, 693 (2003); available online at http://arxiv.org/abs/astro-ph/0302402.
22. D. Spergel *et al.*, *Astrophys. J.*, in press; preprint available online at http://arxiv.org/abs/astro-ph/0302209.
23. S. M. Carroll, "The cosmological constant;" accepted 29 January 2001; available online at www.livingreviews.org/Articles/Volume4/2001-1carroll. S. Carroll has pointed out that the essential feature that leads to acceleration concerns the negative pressure (or tension) rather than the energy, and the vacuum is transparent, not absorbing. We could more properly refer to a universe dominated by dark energy as one controlled by clear tension. I don't think this will catch on.
24. T. Padmanabhan, *Phys. Rep.*, in press; preprint available online at http://arxiv.org/abs/hep-th/0212290.
25. N. Wright, Cosmology Tutorial; revised 13 March 2003; available online at www.astro.ucla.edu/~wright/cosmolog.htm.
26. G. Gamow, *My World Line* (Viking, New York, 1970).
27. A. Einstein, W. deSitter, *Proc. Natl. Acad. Sci. U.S.A.* **18**, 213 (1932).
28. H. Kragh, *Cosmology and Controversy* (Princeton Univ. Press, Princeton, NJ, 1996), p. 54.
29. Y. B. Zel'dovich, *JETP Lett.* **6**, 316 (1967).
30. Y. B. Zel'dovich, *Sov. Phys. Usp.* **11**, 381 (1968).
31. Exact method for placing Λ on the right-hand side of the field equations is available online at http://cfa-www.harvard.edu/~rkirshner/lambda.
32. For a heuristic explanation of the properties of vacuum energy, see the appendix of A. Guth, *The Inflationary Universe* (Addison-Wesley, Reading, MA, 1997), or M. Turner, *Phys. Today* **56**, 10 (2003).
33. $a''/a = -4\pi G/3(\rho + 3P)$, where a is the cosmic scale factor, G is Newton's gravitational constant, ρ is the energy density (with the speed of light, c, set to 1), and P the pressure.
34. A. S. Eddington, *The Expanding Universe* (Cambridge Univ. Press, Cambridge, 1987).
35. More precisely, $\Omega_\Lambda = \rho_\Lambda/\rho_{\mathrm{critical}}$, and $\rho_{\mathrm{critical}} = 3H_o^2/8\pi G$, where H_o is in s^{-1}. The dark energy need not be the cosmological constant, Λ, in which case Ω_Λ would stand in for the density of the dark energy.
36. Supersymmetry can make the energy of the vacuum exactly zero, but that is also not the observed value.
37. P. J. Peebles, B. Ratra, *Astrophys. J.* **325**, L17 (1998).
38. B. Ratra, P. J. Peebles, *Phys. Rev. D* **37**, 3406 (1988).
39. R. R. Calwell, R. Dave, P. J. Steinhardt, *Phys. Rev. Lett.* **80**, 1582 (1998).
40. C. Wetterich, *Nucl. Phys. B* **302**, 668 (1988).
41. R. L. Gilliand, P. E. Nugent, M. M. Phillips, *Astrophys. J.* **521**, 30 (1999).
42. This is described in chapter 11 of R. P. Kirshner, *The Extravagant Universe: Exploding Stars, Dark Energy, and the Accelerating Cosmos* (Princeton Univ. Press, Princeton, NJ, 2002).
43. A. G. Riess *et al.*, *Astrophys. J.* **560**, 49 (2001).
44. M. S. Turner, A. G. Riess, *Astrophys. J.* **569**, 18 (2002).
45. A. G. Riess *et al.*, *Astrophys. J.*, in press.
46. The International Astronomical Union publishes circulars announcing the detection of supernovae. These circulars are available online from http://cfa-www.harvard.edu/iauc/RecentIAUCs.html.
47. S. Carroll, "Dark energy and the preposterous universe," report no. EFI-2001-27, 30 Jul 2001; revised 2 Aug 2001; available online at http://arxiv.org/abs/astro-ph/0107571.
48. P. Garnavich *et al.*, *Astrophys. J.* **509**, 74 (1998); available online at http://arxiv.org/abs/astro-ph/9806396.
49. W. J. Percival *et al.*, *Mon. Not. R. Astron. Soc.* **327**, 1297 (2001).
50. G. Wilson *et al.*, *Astron. J.* **124**, 1258 (2002).
51. S. Carroll, M. Hoffman, M. Trodden, "Can the dark energy equation-of-state parameter w be less than –1?" report nos. EFI-2003-01 and SU-GP-03/1-1, 14 January 2003; revised 4 February 2003; available online at http://arxiv.org/abs/astro-ph/0301273.
52. The ESSENCE website is available at www.ctio.noao.edu/essence/.
53. J. Weller, A. Albrecht, *Phys. Rev. D.* **65**, 3512 (2002).
54. T. Padmanabhan, T. R. Choudhury, "A theoretician's analysis of the supernova data and the limitations in determining the nature of dark energy," 30 Dec 2002; available online at http://arxiv.org/abs/astro-ph/0212573.
55. L. M. Krauss, B. Chaboyer, *Science* **299**, 65 (2003).
56. P. Höflich, J. C. Wheeler, F.-K. Thielemann, *Astrophys. J.* **495**, 617 (1998).
57. F. X. Timmes, E. F. Brown, J. W. Truran, *Astrophys. J.*, in press; preprint available online at http://arxiv.org/abs/astro-ph/0305114.
58. These data were compiled by T. Matheson *et al.*, Center for Astrophysics Spectroscopic Sample, Harvard University, and is available online at http://cfa-www.harvard.edu/oir/Research/supernova.html.
59. More information about the Legacy Survey is available online at www.cfht.hawaii.edu/Science/CFHLS/.
60. W. M. Wood-Vasey *et al.*, *Bull. Am. Astron. Soc.* **201**, 5608 (2002).
61. A. L. Coil *et al.*, *Astrophys. J.* **544**, 111 (2000).
62. S. M. Fall *et al.*, "Science with the Hubble Space Telescope in the next decade: The case for an extended mission"; this report is available online at www.stsci.edu/resources/sm5revise.pdf.
63. S. Perlmutter, *Phys. Today* **56**, 53 (2003). More information on SNAP is available online at http://snap.lbl.gov.

64. Work on supernovae at Harvard University is supported by NSF grants AST 02-05808 and AST 02-06329 and by NASA through the Space Telescope Science Institute grants GO-08641 and GO-09118.