PROBING COSMIC MYSTERIES BY SUPERCOMPUTER

Steady advances in supercomputing hardware and numerical algorithms are beginning to shed light on some of the most recalcitrant problems in astrophysics and cosmology

Michael L. Norman

Astrophysicists have had a love affair with big computers since the dawn of the digital era. Although it is unlikely that the 1946 ENIAC (Electronic Numerical Integrator and Computer) was ever used for astrophysical calculations, Princeton University astronomer Martin Schwarzschild made extensive use of the follow-on MANIAC computer at the Los Alamos laboratory for his pioneering calculations of stellar evolution. Since World War II, simulating the inner workings of nuclear weapons has been one of the key applications driving the development of supercomputer technology. Because many of the same physical processes operate in stars and nebulae as in hydrogen bombs, supercomputers designed with defense needs in mind have been ideally suited to computational astrophysics research. Until the mid-1980s, however, access to supercomputers was limited to a small cadre of researchers at defense laboratories or at a few specialized academic institutions.

The establishment of the National Science Foundation Supercomputing Centers in 1985 opened up access to state-of-the-art supercomputers to the entire academic community. This development, and the subsequent creation of state and regional supercomputing centers, the opening up of NASA and Department of Energy supercomputing facilities, the emergence of powerful and affordable workstations and the growth of the Internet, have all played a role in the hundredfold increase in the ranks of computational astrophysicists. Computational astrophysics research has been enjoying a decade of unprecedented growth and progress.

Today's most powerful supercomputers are a billion times as fast as the ENIAC, and they contain tens of billions of bytes of random-access memory. That's enough to store the position and velocity of every star in a small galaxy. These tremendous strides in hardware performance have been matched by equally impressive advances in algorithmic efficiency. (See, for example, the article by Joshua Barnes and Lars Hernquist in PHYSICS TODAY, March 1993, page 54.) These two trends are multiplicative, with the result that every year in recent decades has seen roughly a doubling of the complexity of the problems that computational astrophysicists can tackle. These astonishing improvements have opened up exciting new vistas in astrophysical modeling, and they have brought us to the threshold of solving some of nature's most perplexing cosmic mysteries.

Broadly speaking, the goal of computational astrophysics is to faithfully simulate from first principles the diverse objects and exotic phenomena of the astronomical universe—quasars, neutron stars, supernovae, star forming regions and the like. Numerical cosmology is computational astrophysics on a global scale. Its grandiose objective is to self-consistently simulate a representative patch of the observable universe by letting it evolve from primordial initial conditions to the present time. The ultimate goal is to understand the complex interplay of forces and processes that govern cosmic phenomena. Another goal is to deduce the detailed evolutionary history of the universe by adjusting cosmological parameters to matching the computed evolution to all the relevant observations. Thus, one seeks to determine key parameters such as the Hubble constant and the mean mass density of the cosmos.

The following section briefly describes why supercomputers are needed to these ends, and how they are used to advance our understanding. Subsequent sections provide examples of significant progress toward the solution of long-standing problems in astrophysics and cosmology. The final section discusses current developments in hardware and numerical algorithms that will underpin future progress.

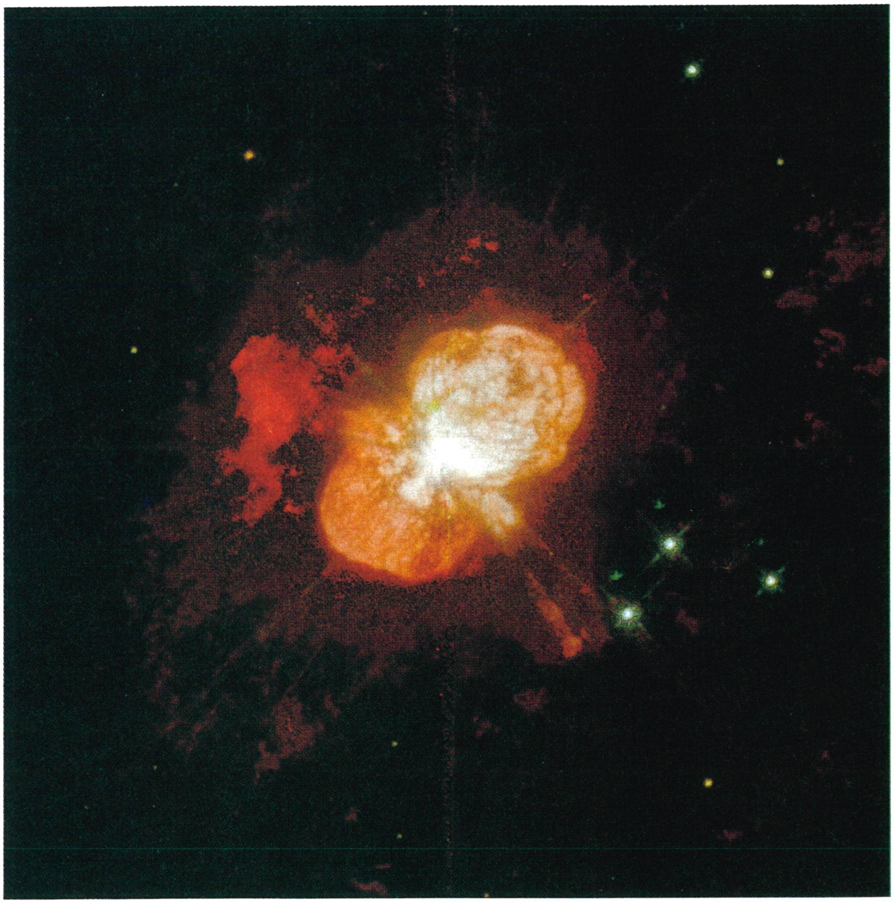
The computer as laboratory and observatory

Astronomy is unique among the physical sciences in that one is permitted to look, but not to touch. In the late 19th century, solar physics pioneer George Ellery Hale was confronted with the problem of interpreting the then mysterious solar spectrum. He argued for the creation of a third branch of astronomical research, which he termed "laboratory astrophysics," to complement observation and theory. In 1906 he wrote: "The immediate imitation in the laboratory, under experimental conditions subject to easy trial, of solar and stellar phenomena, not only tends to clear up obscure points, but prepares the way for developing along logical lines the train of reasoning started by the astronomical works."

In the first half of the 20th century, Hale's vision was fulfilled: The confluence of quantitative spectroscopy, laboratory astrophysics, and atomic and molecular theory gave astronomers some of their most powerful analytic tools, which they used to discover the chemical composition of the stars and nebulae. Additionally, by measuring the Doppler shifts of spectral lines, astronomers could determine the line-of-sight velocity component of an astronomical object. Spectroscopic Doppler measurements have led to some of the most important astronomical discoveries of the twentieth century, including the structure of the Milky Way Galaxy, the expanding universe and quasars.

Computational astrophysics is a kind of laboratory astrophysics in which three-dimensional structure, dynamics and temporal evolution of astronomical systems can be simulated from first principles. This information

MICHAEL NORMAN *is a professor of astronomy and director of the Laboratory for Computational Astrophysics of the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign.*



Eta Carinae, a superluminous star in our Galaxy, illustrates the complexity of astrophysical phenomena. This Hubble telescope image shows twin lobes of gas and dust ejected in 1841. Such gigantic outbursts, though not well understood, can presage supernovae. The thousandfold expansion of computing power expected in the next decades should enable us to model Eta Carinae in detail. (Image courtesy of John Hester and NASA.) **Figure 1**

is generally not directly accessible to observation—typically because evolutionary time scales are too long or because the inner workings are hidden from view. Examples include star formation, stellar convection and interacting binary stars.

Computational astrophysics also is a kind of experimental astrophysics, in which numerical laboratories probe such hidden dimensions. This is done in two basic steps: First, one gets a numerical solution to the equations governing the structure or evolution of the model system. Second, one translates the results of the calculation into observational terms for direct comparison. Computational astrophysics is, if you like, the fourth astronomical methodology, standing synergistically alongside the traditional roles of the observer, the theorist and the laboratory experimenter.

The interplay between computational astrophysics and the other methodologies may need a bit of clarification: Theory interacts with simulation in three essential ways. First, theory provides the mathematical formulation for the numerical model, and it defines the parameter space of solutions to be searched. Second, it incorporates useful analytic properties of the solution (conservation laws, for example) into the numerical algorithms. Such analytic solutions, in fact, provide excellent test problems for validating simulation computer codes. Failure to reproduce an analytic result often stimulates the critical thinking required for inventing more accurate algorithms. Finally, when analyzing the results of a numerical simulation, especially a simulation involving many complex physical processes, one attempts to construct simplifying models that nonetheless capture the essential physics.

For their part, simulations provide realizations of theoretical models that are too complex to be solved analytically. These realizations are in essence the laboratory data that test the theoretical models. In astrophysics, one is often not sure that all of the relevant physics has been included in the model. A simulation's failure to reproduce the observations may indicate missing physics, bad numerics or bad observations. Furthermore, simulations build physical intuition by providing the modeler with direct experience of the complex phenomena embodied in the governing equations. Finally, of course, the simulation must confront the observations.

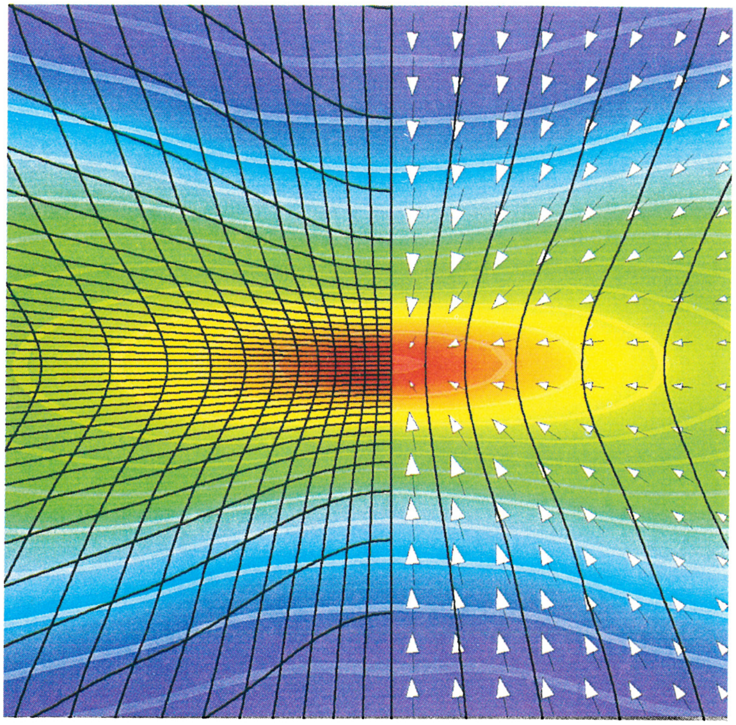
The impressive success of the theory of stellar structure and evolution is a case in point. Without the computational means to solve the structure equations to a high degree of precision, the field would not be anywhere near where it is today. The following are three illustrative examples of fundamental problems in astronomy and cosmology that are inherently multidimensional and involve physics of different kinds on different scales. Thus, they stretch current computing capabilities to the limit. Each example is a topic of fundamental importance that is beginning to yield up its secrets to our vastly improved ability to compute.

Star formation

We consider first the mystery of present-day star formation in the Milky Way. The existence of massive stars such as Eta Carinae (see figure 1), whose core hydrogen fusion lifetimes are less than a thousandth the age of the Milky Way, is proof that star formation is an ongoing process in places such as the Orion Nebula. A complete theory of star formation, which we still lack, must explain the conditions under which it occurs, its rate and the resulting distribution of stellar masses.

We know that stars form in gigantic, cold interstellar molecular clouds found in the spiral arms of our Galaxy. In addition to H_2, these giant molecular clouds, with temperatures on the order of 20 K, contain CO, CN, H_2O and other molecules whose rotational transition lines can be detected at radio wavelengths. With a density on the order of a thousand molecules per cubic centimeter, a giant cloud typically has a mass a million times that of the Sun.

COLLAPSE OF A PROTOSTELLAR CORE in an axisymmetric simulation of a magnetized molecular cloud.[4] While ionized matter is supported by flux lines, neutral matter is gravitationally pulled into the center. Colors and pale contour lines indicate the density of neutral matter, increasing by five orders of magnitude from periphery to center, as its velocity (indicated by arrows) decreases. Black lines (right) are magnetic field lines, and the black grid (left) shows the adaptive coordinate mesh used for this simulation. (Courtesy Robert Fiedler and NCSA.) FIGURE 2



Giant molecular clouds are rather lumpy, containing cores that are denser and cooler than their surroundings. These molecular cores are believed to be sites of gravitational contraction and incipient star formation. A proper theory must explain the origin of dense molecular cores and predict the sequence that transforms them into stars.

A combined assault of theoretical analysis and supercomputer simulations is beginning to reveal how this process works. James Jeans formulated the classical picture of star formation, early in this century. A key result of his analysis is the concept of the Jeans mass,

$$M_{\mathrm{J}} = \left(\frac{5kT}{G\mu}\right)^{3/2} \left(\frac{3}{4\pi\rho}\right)^{1/2}$$

below which a uniform, spherical gas cloud of temperature T, density ρ and molecular mass μ would be in stable gravitational equilibrium. If a cloud grows beyond the Jeans mass, gravity overwhelms thermal pressure and collapse ensues, continuing as long as the gravitational energy released by contraction can radiate away on a time scale short compared to the collapse time. This is the case for gas densities below about 10^{10} molecules per cubic centimeter. To good approximation, the collapse proceeds isothermally, because the gas temperature is set by a balance between cosmic-ray heating and molecular radiation cooling. Therefore, the cloud's thermal energy remains constant while its gravitational potential is released. The result is a runaway collapse in which the gas quickly accelerates to freefall velocities.

In a highly simplified model, the cloud would collapse to infinite density in the so-called free-fall time,

$$t_{\mathrm{f}} = \left(\frac{3\pi}{32G\rho}\right)^{1/2}$$

But in more realistic models, the central region of the cloud eventually becomes opaque to its cooling radiation, and its thermal energy increases until the core is supported against further collapse by its pressure.

Spherically symmetric hydrodynamical simulations by Richard Larson and others in the late 1960s showed that the opaque core will contract quasistatically and heat up until it gets hot enough to dissociate molecular hydrogen. This endothermic phase transition robs energy from the core, causing it to enter a second collapse phase. The scenario repeats itself as changes in the opacity and equation of state play havoc with the core's energy and pressure balance, until one finally gets a protostar—still without thermonuclear fusion.

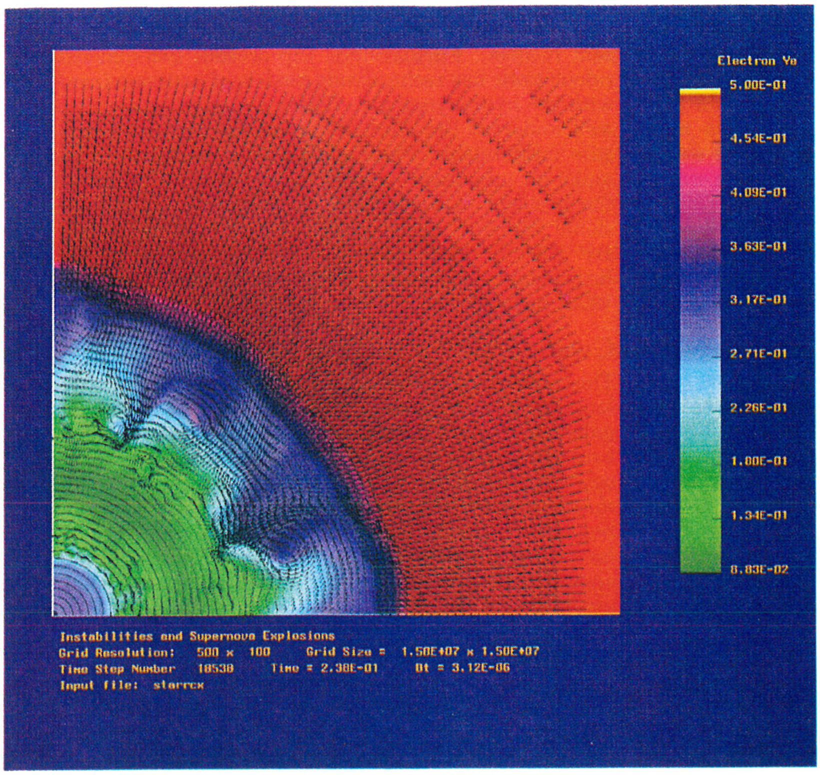
According to these models, the remainder of the cloud mass then rains down onto the protostar on the free-fall time scale of the initial cloud. As the protostar accumulates this additional mass, it contracts quasistatically on a radiation–diffusion time scale until its central temperature becomes sufficient to fuse hydrogen and a true star is born.

This classical scenario suffers from a number of well-known problems. First there is the efficiency problem: According to the classical picture, one would expect a giant molecular cloud to fragment into gravitationally unstable clouds of Jeans mass—roughly 10 solar masses—which would then collapse and form stars in about a million years. Even if each collapsing cloud produced only one star, a giant molecular cloud would transform itself into tens of thousands of new stars in a million years. But we know that the rate of star formation is several orders of magnitude slower than that. What's holding the molecular clouds up?

A related problem is how to account for the existence of dense molecular cores. Though they are known to have masses far above the Jeans mass, they show no evidence for gravitational collapse. Then there's the angular momentum problem: An interstellar cloud of one solar mass has roughly 10^4 times as much angular momentum as the Sun. Without an efficient means of shedding angular momentum, a collapsing cloud would become a centrifugally supported disk long before it reaches stellar density.

A new picture of star formation, first explored by Leon Mestel and Lyman Spitzer[1] and George Field in the 1950s and later developed by Telemachos Mouschovias,[2] solves these problems by considering the dynamical effects of the interstellar magnetic field. Mouschovias and Spitzer[3] showed that a cloud of mass M can be magnetically supported against collapse by a magnetic flux Φ if Φ/M exceeds $(63\,G)^{1/2}$. From observed Zeeman splitting of molecular lines one finds a typical magnetic field of 30 microgauss. A frozen-in field of that magnitude can increase the critical mass for gravitational instability a hundredfold.

But how can a magnetically supported cloud form stars? As neutral atoms slip through the magnetic field, their inward migration is retarded only infrequently by



CONVECTIVE INSTABILITIES in an axisymmetric simulation of the collapsed iron core of a 15-solar-mass star just 70 milliseconds before supernova explosion. Neutrinos from electron capture in the core heat the surrounding matter to produce vigorous convection that triggers the explosion. Colors indicate surviving electron population per nucleon and arrows show velocities. The red/purple discontinuity indicates an accretion shock that liberates protons from heavy nuclei. There the shock stalls briefly, to be reenergized by convective eddies in the final 70 ms. (Courtesy of Adam Burrows, John Hayes and Bruce Fryxell, University of Arizona.) FIGURE 3

collisions with ions strongly coupled to the magnetic field. This process, known as ambipolar diffusion, proceeds on a time scale much longer than the free-fall time. Thus, dense cores contract quasistatically until Φ/M in their central flux tubes falls below the critical support value and the core begins to collapse inside the magnetically supported cloud envelope. In addition to the longer time scale, the efficiency of star formation is further reduced by the fact that only the central region of the cloud collapses. The angular momentum is removed from the cloud by torsional Alfvén waves, through a process known as magnetic braking.

To study the magnetic star-formation process in detail, one must solve the time-dependent equations of multifluid magnetohydrodynamics in two or three dimensions, taking account of self-gravitation, rotation and radiative transfer. Such calculations require supercomputers because of the problem's inherent size and complexity. Figure 2 shows a calculation of the collapse of a molecular cloud core initially supported by its magnetic field, carried out by Mouschovias and Robert Fiedler at our University of Illinois supercomputing center. The calculation ignores the effects of rotation; it models only the isothermal phase of the evolution.[4] Still the problem remains very difficult because of the large range of length and time scales that must be considered. Fiedler and Mouschovias used an adaptive grid to resolve the protostellar core, whose diameter is ten thousand times smaller than the entire cloud. The inclusion of rotational effects in a subsequent calculation has confirmed the effectiveness of magnetic braking.

More powerful supercomputers will be required to follow the core's collapse to stellar densities, and the dynamical equations will have to be supplemented with additional equations for the energy of the gas and the radiation field. An even more ambitious undertaking would be to model a large piece of a nonspherical molecular cloud in three dimensions, to study the generation, propagation and damping of magnetohydrodynamic waves, as well as their influence on core formation and collapse.

Core collapse supernovae

A star of ten or more solar masses ends its life with a spectacular supernova explosion so luminous that it briefly outshines an entire galaxy. One of the major challenges for astrophysical modeling in recent decades has been to understand how the explosion mechanism works. Observations tell us that the explosion energy is of order 10^{51} ergs. But because the observations probe the mechanism only indirectly, progress has relied almost entirely on theoretical and numerical models. Despite general agreement on the basic picture, increasingly sophisticated numerical models have shown the physics to be astonishingly complex. Whether or not a model yields an explosion turns out to be highly sensitive to the physics input and the approximations.

The detection of neutrinos from supernova 1987A confirmed the basic picture discovered by the numerical models. Such "type II" supernovae are thought to originate from the collapse of the core of a massive star once its fusion energy sources are exhausted. Owing to a succession of nuclear burning stages, such a star has a layered, onionlike structure, with hydrogen on the outside and iron, with the tightest binding of all nuclei, at the center. The iron core becomes unstable and collapses because of electron capture, which robs the core of the electron degeneracy pressure that had supported it.

Electron capture reactions continue during the collapse until the increasing density prevents the resulting neutrinos from escaping the core on the millisecond collapse time scale. Eventually neutrino production, together with the rapid density increase, causes the neutrinos to become degenerate. But still the pressure is insufficient to halt the collapse until supernuclear densities are reached and nucleon degeneracy and repulsion become dominant.

When the collapse is finally halted, the inner core rebounds, forming a shock wave that propagates out into the still-infalling outer core and dissociates its heavy nuclei. In the very dense matter below the shock wave, inverse beta decay turns protons into neutrons, forming a proto–neutron star in a few seconds.

A now-disproved, purely hydrodynamic theory asserted that this shock front advances through the infalling stellar envelope, reverses its motion and eventually ejects the envelope. But the 8.8 MeV per baryon required to dissociate nuclei, together with neutrino losses from the high-temperature shocked material, would weaken and eventually stall the shock wave, in the absence of some reenergizing mechanism, before it could blow off the star's outer mantle.

In the mid-1980s Hans Bethe and James Wilson[5] discovered an alternative "late-time" mechanism by availing themselves of a powerful supercomputer at the Lawrence Livermore National Laboratory to add more physics to the simulations and running them to much later times. In the late-time mechanism, the stalled shock is revived by tapping into the enormous neutrino flux emitted by the proto–neutron star. During collapse, the released gravitational energy is comparable to the binding energy of a neutron star, about 3×10^{53} ergs. Most of that energy is radiated away as neutrinos. Less than 1% of the neutrino energy deposited behind the stalled shock would be enough to power the supernova explosion. But that's easier said than done. Inside the high-density proto–neutron-star core the mean free path for neutrinos is short. Thus, the neutrinos behave diffusively. But as the neutrinos diffuse out of the core, their mean free path becomes longer and eventually reaches the free-streaming limit at which the neutrinos scarcely interact with matter. Determining how much the neutrinos contribute to reenergizing the stalled shock wave before they start free-streaming requires a very accurate treatment of neutrino interaction and transport.

Spherically symmetrical hydrodynamic simulations including these neutrino effects were carried out by Wilson[6] in the early 1980s. They produced explosions by means of neutrino heating at late times—about half a second after bounce. But simulations by others failed to reproduce these provocative results. They found the mechanism to be extremely sensitive to the details of neutrino transport and choice of nuclear equation of state, which is uncertain at supernuclear densities. For example, because neutrino opacities increase rapidly with energy, one has to solve for the neutrino spectrum at every radial point, making the problem effectively two-dimensional despite the spherical symmetry of the model.

Other researchers explored asymmetric models of core collapse. Although it was known that the core had unstable entropy and composition gradients, the convective transport had nonetheless been calculated, for simplicity, in spherically symmetric models using phenomenological mixing lengths. Then in 1992, Marc Herant, Willy Benz and Stirling Colgate[7] carried out a two-dimensional calculation (on a workstation!) that showed convective motion to be a surprisingly efficient way of transporting energy. They found that convective eddies were larger than expected, and that dissociated matter dredged up from the bottoms of these eddies re-associated into nuclei, releasing nuclear energy behind the shock wave and thus assuring that the supernova explosion would not fizzle. Others soon confirmed this result,[8] and the race was on to refine such calculations. Figure 3 shows a recent, higher-resolution calculation done by Adam Burrows and University of Arizona colleagues on a Cray at the Pittsburgh supercomputing center.

If convection is indeed the key piece of physics for type II supernovae, which at this point is by no means established, then we will be keeping supercomputers busy for years to come, trying to assess its importance. That's because turbulent convection is inherently a three-dimensional phenomenon, whereas current models are restricted to two dimensions. It is also a multiscale phenomenona, requiring numerical resolutions not yet achieved in present models. Furthermore, the present two-dimensional calculations use rather crude models of neutrino transport with the hydrodynamics. Better neutrino transport could well undercut the effectiveness of the Herant–Benz–Colgate mechanism. Finally, if nuclear reactions do indeed make an important contribution to heating the gas, then the problem becomes more difficult still. Three-dimensional computer simulations of turbulent reactive flows are still in their infancy. (See the article by George Karniadakis and Steven Orszag in PHYSICS TODAY, March 1993, page 34.)

Cosmological structure formation

The universe exhibits structure on a vast hierarchy of length and mass scales, from planets to superclusters of galaxies. Explaining the origin and evolution of this structure over cosmic time is the province of numerical cosmology. As Douglas Adams observed in *The Restaurant at the Edge of the Universe*, "The universe is an unsettlingly big place." So it should come as no surprise that simulating it requires a very big computer. Any model realistic enough to confront the great array of observations now available must span a vast range of mass and length scales, and it must incorporate a lot of detailed physics of radiation and gases.

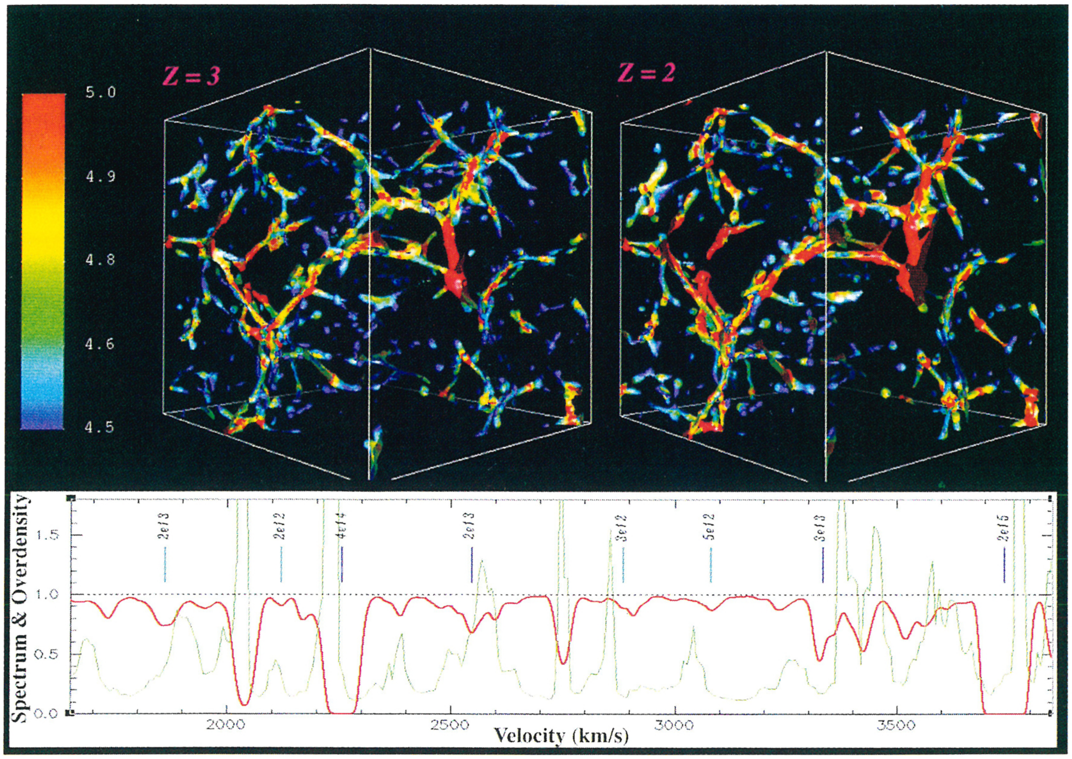
The basic scenario of cosmological structure formation is as follows: Small primordial density fluctuations are amplified by gravity as the universe expands and cools. From the COBE (Cosmic Background Explorer) satellite measurements of the cosmic microwave background anisotropies, we know that the amplitude of the density fluctuations in the primordial gas when matter first became transparent (3×10^5 years after the Big Bang) was only a few parts in a million. (See PHYSICS TODAY, June 1992, page 17.) Because of the Hubble expansion, gravitational instability growth would have been linear rather than exponential as long as the fractional density fluctuation amplitude $\delta\rho/\rho$ was small.

If we parametrize time by the cosmological redshift z, the scale length of the expanding universe grows like $1/(1+z)$, and so, in that early linear epoch, did the density fluctuation amplitudes. Today, $z = 0$; it was about 1300 when the primordial gas first became electrically neutral and hence transparent. Thus the gravity of the gas alone is insufficient to have formed highly nonlinear structures such as galaxies, for which $\delta\rho/\rho > 1$.

That's why cosmologists posit the existence of "cold dark matter" (CDM) that would have decoupled from the radiation field much earlier than the gas did. Such cold dark matter, being impervious to electromagnetism, would have begun clustering long before the ordinary gas. The growing dark matter perturbations would have formed potential wells for the gas to pool in. By about $z = 100$, the gas fluctuations would have caught up with the dark matter, producing fluctuation amplitudes large enough to go nonlinear by $z = 5$, when we need them to form galaxies and quasars.

It requires numerical simulations to follow the details of how structure formation proceeds in the nonlinear regime. The development of the cold-dark-matter theory in the mid-1980s coincided with the rise of minicomputers and the National Science Foundation supercomputing centers. So the late 1980s witnessed a vigorous growth industry in numerical cosmology.

Because dark matter is believed, on a variety of observational grounds, to contribute most—perhaps 95%—of the total mass density of the universe (see PHYSICS TODAY, August, page 17), the first simulations ignored the dynamics of the primordial gas altogether, concentrating



VOIDS AND FILAMENTS in the cosmic gas distribution at two different redshifts Z, evolved from primordial gas, radiation and cold dark matter in a simulation of a cosmic cube 10 megaparsecs on a side. Colors indicates the temperature scale; red is hottest. The bottom panel shows a Lyman-alpha forest spectrum (red curve) generated from these simulations. The redshift of each Ly-α feature is proportional to the recessional velocity (abscissa) of the responsible absorbing hydrogen. The green curve indicates the corresponding redshift distribution of the overall baryon density. The blue ticks (labeled by hydrogen column density in atoms per cm^2) mark weak "Lyman alpha underbrush" lines generated by these simulations. (Adapted from ref. 16.) FIGURE 4

instead on simulating the collisionless dynamics of self-gravitating dark matter particles by means of N-body techniques. These simulations integrate the equations of motion for N point masses in a reference frame comoving with the expanding universe. The calculations are performed in a periodic cubic domain large enough to encompass hundreds or even thousands of galaxies.

The results depend on several free parameters, including the mean mass density of the universe and the Hubble constant, H_0, and also on a free function $P(k)$, the initial Fourier power spectrum of density fluctuations. The form of this fluctuation power spectrum is fixed by theory; only its amplitude is unspecified.[9] Through detailed comparisons between the numerical simulations and observations, cosmologists hope to determine $P(k)$, together with H_0 and the mass density.

The enormous dynamic range requirements of cosmological simulations stem from the need to handle large statistical samples of objects such as galaxies, while at the same time resolving their internal structures. That will require multiscale algorithms that span ranges of at least 10^4 in length and 10^9 in mass. Simulations approaching these dynamic ranges are now feasible with parallel computers and fast N-body algorithms. Such studies of the nonlinear dynamics of structure formation have already shown that cold dark matter clusters hierarchically, with subgalactic mass scales collapsing first, followed by galaxy creation and finally the formation of large clusters.[10]

This so-called bottom-up theory is indeed consistent with observations. A major challenge to the CDM theory came with the discovery of large-scale cosmological structure. We now know that galaxies are concentrated into enormous sheets and filaments surrounding nearly empty circular voids.[11] CDM models reproduce this spongelike structure qualitatively, but not quantitatively. The COBE measurements fix the amplitude of the power spectrum on very large scales. If one uses that value to normalize the entire CDM power spectrum, the resulting simulations tend to overproduce galaxy clusters by an order of magnitude.[12] That's why some cosmologists have been heard to declare that "CDM is dead."

Two avenues are being pursued to resolve this discrepancy. First, one can construct alternative fluctuation power spectra explicitly to match the observations. That corresponds to making alternative assumptions about the composition of the dark matter. "Cold" means that the particles, being heavy, are slow enough to cluster easily. "Hot" dark matter, by contrast, would consist of very light particles that would evade clustering because of their relativistic velocities. The now popular hybrid cold + hot dark matter (CHDM) model assumes an admixture of cold dark matter and low-mass neutrinos. Also under study are open-universe models (with mass densities too small for gravitational closure) and models with a nonzero cosmological constant that serves to retard gravitational clustering.

The second avenue incorporates into the simulations detailed gas and radiation physics, glossed over by the pure dark-matter simulations, in order to provide a more realistic description of how galaxies form. But incorporating all that physics substantially increases the complexity of the calculations.

Despite our most powerful supercomputers, galaxy formation remains an important unsolved problem. Reliable calculations from first principles are still out of reach. That's because the large-scale simulations cannot resolve the scales on which stars form in protogalaxies. Nonetheless, simulations with dark matter and simple gas physics have succeeded in producing flattened, rotationally supported gas disks, looking very much like galaxies, embedded in dark-matter halos.[13]

Very recently and somewhat unexpectedly, hydrodynamic cosmological simulations have scored an apparent success at explaining the origin of the so-called Lyman-alpha forest, a mystery that has persisted since its discovery in 1971. The Lyman-alpha forest is a thicket of hydrogen Ly-α absorption lines seen in the spectra of distant quasars. The lines are caused by absorption in the intervening intergalactic medium along the line of sight to the quasar. Each line is redshifted by the Hubble recessional velocity of the particular intergalactic cloud whose neutral atomic hydrogen absorbs the quasar's light. A typical quasar spectrum contains about a hundred of these redshifted Ly-α absorption lines, and there are roughly a thousand known quasars. The resulting 10^5 redshifted Ly-α lines represent the largest cosmological database we have; but it is only now being assembled. Combined with the right theory, it will tell us about the detailed structure of the intergalactic medium before and during the epoch of galaxy formation.

Early models of the forest envisioned discrete gas clouds, perhaps protogalaxies, embedded in a less dense medium. Recently three groups supported by an NSF-funded computational grand-challenge project in cosmology showed that the observations were a natural consequence of hierarchical structure formation. In 1994, Jeremiah Ostriker, Renyue Cen and coworkers at Princeton showed that the features of the Ly-α forest can be attributed to the filamentary structure of the intergalactic medium.[14] A 1995 simulation by Lars Hernquist and colleagues at the University of California, Santa Cruz, showed that the power-law distribution of observed absorption column densities was well described by their model of cold dark matter and gas.[15] Soon thereafter simulations by my group at the University of Illinois were able to reproduce the weakest Ly-α absorption lines (sometimes called the Lyman-alpha underbrush) observed by the Keck telescope and show that they originate in "minivoids" much smaller than the enormous voids one sees in the distrbution of galaxies.[16] (See figure 4.)

Teraflops and beyond

Hardware performance is expected to continue improving at its wonted rate, with computing speed doubling roughly every 18 months, for at least another decade. Supercomputers with peak speeds in excess of a teraflop (10^{12} floating-point operations per second) will be operational at the Department of Energy weapons labs by 1998. Ten years after that, we can expect to have 100 teraflops.

In academia, the NSF supercomputing centers plan to have machines capable of sustained teraflop performance by the turn of the century. Computer memory is likely to keep pace. These machines will be about a thousand times as powerful as those employed for the work described in this article. This extraordinary progress, expected from the relentless advances in microprocessor technology and massive parallelism, will allow astrophysicists to construct models of realistic complexity.

For example, a thousandfold improvement is more than sufficient to generalize the star-formation and supernova models to three dimensions and incorporate radiative transfer. In cosmology, future hardware advances will let simulations take account of more physical processes over a greater range of spatial scales. But raw computer power alone will not be enough. For example, the simplest algorithms for evolving astrophysical fluids sample continuous field variables such as density and velocity on a uniform grid. The simulation advances these now discrete variables in time by extrapolation. The dynamic range of the simulation is increased by using more grid points N. But then the memory requirement scales as N^3 and the number of operations required scales as N^4. Thus, a thousandfold increase in computer speed yields less than a sixfold improvement in dynamic range for a given running time. So we will need smarter algorithms to go with the hardware advances, if we are to confront the enormous dynamic ranges present in astrophysical systems.

One algorithm, called smoothed particle hydrodynamics, replaces static grid points with variable-sized Lagrangian fluid elements that automatically track and rescale with the flow. (See the article by Barnes and Hernquist in PHYSICS TODAY, March 1993). Another promising approach, which my students and I are exploring, employs an adaptive, multilevel grid hierarchy that automatically refines itself to maintain adequate spatial resolution in select regions. Applying this algorithm to a simulation of galaxy cluster formation, we use adaptive-mesh refinement to achieve very high spatial dynamic range in regions where galaxies are forming while we make do with much lower resolution in diffuse intergalactic regions where very little structure is present. An equivalent calculation using a uniform grid would require 4000 times the memory and 20 000 times the CPU performance of the computer we use.

Harnessing all these advances will require more powerful programming languages. To that end, NSF, NASA and DOE have created programs to fund interdisciplinary teams of application scientists, computer scientists and computer technicians to tackle so-called grand challenge problems. In the astronomical disciplines, these teams are creating powerful new computer simulations of stellar convection, galaxy formation and the evolution of binary neutron-star and black-hole systems. Many of the tools developed in these projects will be applicable to other, still-elusive problems such as the formation of planetary systems and the nature of quasars.

References

1. L. Mestel, L. Spitzer Jr, Mon. Not. R. Astron. Soc. **116**, 503 (1956).
2. T. Mouschovias, in *Physics of the Interstellar and Intergalactic Media*, A. Ferrara, C. McKee, C. Heiles, P. Shapiro, eds., ASP Conf. Series Vol. 80, p. 184, Astron. Soc. of the Pacific, San Francisco (1995).
3. T. Mouschovias, L. Spitzer Jr, Astrophys. J. **210**, 326 (1976).
4. R. Fiedler, T. Mouschovias, Astrophys. J. **415**, 680 (1993).
5. H. Bethe, J. Wilson, Astrophys. J. **295**, 14 (1985).
6. J. Wilson, in *Numerical Astrophysics*, J. Centrella, J. LeBlanc, R. Bowers, eds., Jones & Bartlett, Boston (1985), p. 422.
7. M. Herant, W. Benz, S. Colgate, Astrophys. J. **395**, 642 (1992).
8. A. Burrows, B. Fryxell, Science **258**, 430 (1992). H. Janka, Astron. Astrophys. **256**, 452 (1992).
9. J. Ostriker, Ann. Rev. Astron. Astrophys. **31**, 689 (1993).
10. E. Bertschinger, J. Gelb, Comput. Phys. **5**, 164 (1991).
11. S. Gregory, L. Thompson, Astrophys. J. **222**, 784 (1978). M. Geller, J. Huchra, Science **246**, 897 (1989).
12. M. Davis, G. Efstathiou, C. Frenk, S. White, Nature **356**, 489 (1992).
13. A. Evrard, F. Summers, M. Davis, Astrophys. J. **422**, 11 (1994).
14. R. Cen, J. Miralda-Escude, J. Ostriker, M. Rauch, Astrophys. J. Lett. **437**, 9 (1994).
15. L. Hernquist, N. Katz, D. Weinberg, J. Miralda-Escude, Astrophys J. Lett. **457**, 51 (1996).
16. Y. Zhang, A. Mieksin, P. Anninos, M. Norman, Univ. of Illinois preprint astro-ph/9601130 (1996). ■