Chapter 8
Extraterrestrial Computing: Exploring the Universe with a Supercomputer

Larry L. Smarr*

1 The Bahcall Committee Report–A Decade of Discovery

I've been aided in preparing my talk by a process which the astronomical community undergoes every ten years. This formal process, under the National Academy of Sciences and the National Research Council, has the community debate the future of all our major subfields and to make recommendations on the priorities for funding in both space and ground based astronomy for the next ten years. Dr. John Bahcall, from the Institute for Advanced Study in Princeton, headed the Astronomy & Astrophysics Survey Committee which has recently produced the Academy document "A Decade of Discovery." This study was divided into a number of panels, such as Optical/Infrared Observing, Radio Observing, Planetary Astronomy, etc. Each panel consisted of several dozen of the most imminent researchers in the country in that subfield. They spent a year brainstorming together—meeting, writing, and arguing—to set priorities for their subfield of astronomy and astrophysics.

I was chosen to be the head of the computing panel, which comprises both data processing and simulation. Our field has been involved with supercomputers for forty years. It is claimed that roughly half of John von Neumann's Maniac supercomputer was used by Princeton University astronomer Martin Schwarzchild during the early 1950s to compute stellar evolution models. Thus, computational astrophysics has been around as long as the von Neumann architecture. Our field simply wouldn't exist without supercomputers.

One big change from the preceding report on astronomy and astrophysics of ten years ago is that supercomputing was not a major concern then and computational astronomy was hardly mentioned. This time computing has

*Director, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, Illinois.

been highlighted by being chosen to have its own chapter in the Bahcall Committee report.

This dramatic shift during the last decade reflects how quickly High Performance Computing and Communications is changing the manner in which we perform scientific research. What is happening to astronomy and astrophysics, as is the case with most sciences, is the transformation of the science to a digital form. In the 1990s, a national information infrastructure to support digital science will arise, namely the gigabit per second NREN, the National Research and Education Network, which will hook together supercomputers, massive data archives, observational and experimental instruments, and millions of desktop computers. In short, we are becoming an electronic scientific community.

During the next decade all aspects of our field will be involved in this fundamental change. Telescopes will be transformed into supertelescopes by coupling supercomputers and image processing devices to the steel and glass. Theoretical astrophysics will be done in a numerical laboratories, some of which can support large simulations in which you assume that your basic data cube is measured in tens of gigabytes. A distributed digital library will become essential for research, because the observational data will be a part of this library as well as the results of theoretical simulations and the scientific literature. Observing will be possible in both real-time and remote modes. Software standards in community software both for simulation and observation will become necessary for the progress of the field.

I will only have space to cover a few of the Grand Challenges of computational astronomy and astrophysics, so I will take the liberty of choosing them from areas I have personally worked in. The topics I will briefly cover, all of which will require very large scale computation in the 21st century, will include the structure of the universe, extragalactic jets, radio supermap observations, accretion onto supermassive black holes, and gravitational radiation from black holes. I won't have time to cover many other equally fascinating subjects such as simulations of star formation and the interstellar medium, dynamics of stars and stellar atmospheres, supernovae, or formation and evolution of galaxies. Neither will I be able to cover the broad revolution in observational astronomy which includes the entire electromagnetic spectrum.

2 Large Scale Structure of the Universe

Let's start at the beginning of the universe. Colorplate 7 shows a scientific visualization NCSA did for ABC Television's 20/20 for their show on Stephen Hawking. It's simply a modern visualization of an old idea for representing the expansion of the universe. Instead of trying to represent all

of three-dimensional space, we think of a two-dimensional surface with dots representing individual galaxies. It is convenient to think of this two-surface as the surface of a balloon because, according to Einstein, space acts like an elastic medium. It can be curved, stretched or rippled with these deformations representing the cosmology, black holes, gravitational radiation of the theory.

If you sit at one of the points on the surface (representing our galaxy) and look out in all directions on the surface of the balloon as it is blown up, you will observe that all the other points are moving away from you. If you sit at a different point in space, you will find that again all the other points are moving away from you. This underlying expansion of space is why distant galaxies appear to be moving away from us. As a measure of that motion, the spectrum of light from a galaxy has spectral lines which are red shifted an amount proportional to their distance from us. This simple formula is termed the Hubble relationship.

Since the further away a galaxy is, the earlier in the history of the universe its light left to get here, we can also conclude that the universe was expanding faster in the past than it is in the present. Any textbook on astronomy will describe this expansion with a concise analytic formula. This simple view of a homogeneous cosmology dominated research from the time that Einstein's theory of general relativity was published in 1916 until the 1980s. As we will see, Nature has conspired to be quite different and the true complexity of our cosmology definitely requires supercomputers to decipher.

Individual galaxies come in two types: spiral and elliptical. The spiral galaxies are flat disks of young stars, while the elliptical galaxies are ellipsoids of old stars. Like many things in Nature this clustering of matter in the universe is hierarchical. Not only are the stars clustered in galaxies, the galaxies in clusters of galaxies, but the clusters are organized into superclusters.

Astronomers have made catalogs of the observed galaxies and clusters of galaxies over the years. Each cluster has a red shift and therefore, by the earlier mentioned Hubble relationship, an inferred distance. Several years ago NCSA worked with R. Brent Tully, a professor at the University of Hawaii, to take one such catalog and create a three-dimensional scientific visualization of the database. The colored surfaces shown in Colorplate 8, created with Wavefront software, represents the local density of galaxies per cubic megaparsec (a parsec is equal to 2.36 light years of distance). This technique lets you visually explore how the visible matter in the universe is structured.

What you see is the galaxies don't just appear uniformly throughout space but their distribution is very clumpy. Our galaxy is at the center of the sphere. The blank wedges denote where we have to look outward through the plane of our own galaxy, with the result that the dust in the galactic

plane obscures our view so that we don't see any galaxies there.

For this digital library entry to be of use for comparison with supercomputer cosmological simulations, we need to refine the granularity of the observational data. What is needed is thousands of spectra of individual galaxies in a large sample of space. This is precisely what the Harvard–Smithsonian Redshift Project has been involved with for many years. This systematic work by Margaret Geller and John Huchra has already led to a refinement of the clumpy picture of the structure of visible matter in the universe—a picture which seems to resemble a froth of soap bubbles. Most of the volume of space is fairly empty of galaxies (the "Voids"), with the majority of galaxies found clustered into large curved sheets (such as "the Great Wall". Dr. Geller is now beginning a collaboration with NCSA in which we will create a variety of three-dimensional visualizations of the observed structure of the nearby universe.

Why does the visible universe have this highly inhomogeneous structure? Astrophysicists have concluded that the basic reason for this clumpiness exhibited by the visible matter is that it only accounts for 10 percent of the mass in the universe. The other 90 percent is an unknown form of matter, presumably some exotic type of elementary particle matter left over from the early stage of the universe. If we assume that dark matter is very clumped, then the visible matter would just be gravitationally pulled into the denser regions of dark matter and provide an illuminating tracer of the underlying gravitational potential. In fact, it has been known since the work of Zwicky in the 1930s that clusters of galaxies cannot hold themselves together solely by the gravitational pull of the cluster galaxies on each other, but there must be dark matter permeating the cluster and providing most of the gravitational mass. More recent observational work by a wide range of researchers has convinced us that this lumpiness of the dark matter occurs on scales larger than galaxy clusters as well.

For a number of years now, theorists have realized that one can use supercomputers to simulate the formation of large scale structure in the dark matter. One assumes that the universe contains dark matter which has started at an early stage with a slight perturbation from homogeneity. As the universe expands the dark matter becomes more and more inhomogeneous, until finally it becomes highly fragmented into pancakes and filaments. Over five years ago Joan Centrella and Adrian Melott performed such a supercomputer simulation. Joan then worked with Craig Upson of Digital Productions to visualize the data. Craig worked on a Cray Research supercomputer to create an animation of the density contours of the dark matter in a cube of comoving volume. Colorplate 9 is a still from that animation, showing the collapsed dark matter forming into filaments and pancakes. The blue surfaces are contours of lower density and the red of higher density.

These visualizations show a structure similar to that presumably being

illuminated by the galaxies in Nature. However, a grand challenge for the 21st century is to simulate the visible matter gravitationally coupled to the dark matter and see if that system does end up with galaxies and clusters of galaxies distributed as the observers report them to be. This means computing the evolution of visible matter, which is a gas composed primarily of hydrogen and helium and which uniformly filled the universe early on. It is not collisionless, of course, and so it shocks and heats up as it flows into the forming gravitational potentials of the evolving structure of the dark matter and eventually in some order forms stars, galaxies, and clusters of galaxies.

So what is needed here is a complete code to follow this coupled system. That is you have the dark matter, which is being evolved by particle-in-cell or cloud-in-cell techniques. You then have full gas dynamics in a gravitational field which requires solving the Euler equations coupled with electromagnetic radiation along with a Poisson solver for gravitation. What you want to be able to see is the dark matter collapsing into pancakes and filaments, while the gas falls into those valleys and fragments, shocks, heats and then cools by electromagnetic radiation. This would be fairly straightforward if it were not for the problem of multiple length scales.

One needs spatial resolution from the scale of a galaxy to the scale of the large scale structure, which is roughly one thousand times larger. The brute force way would be to use a uniform grid on a cube a thousand on a side or with one billion grid zones. This leads to a memory requirement of tens of gigabytes. The evolution would require roughly a gigaflop-year worth of computing. A less brute force manner in which to do this computation would be to include adaptive mesh refinement and/or domain decomposition. This would reduce the number of grid zones and memory size at the expense of more complex program logic. The execution time might be reduced or increased as a result. Michael Norman (research scientist at NCSA) and his collaborators are beginning to pursue the creation of such a community computer code on the Connection Machine.

This is a well-posed problem; it has a well-posed accurate solution. What I want as a scientist is the solution, then I want the visualization and interactive exploration tools to get into that universe and live in the solution in sufficient detail with sufficient interactivity that I intimately know all of its characteristics. I can then compare the simulated universe with data from the real universe discussed above. These comparisons will show whether different statistics of different types of elementary particles (defining different types of dark matter) lead to a universe that's closer to the observed universe or not. This is certainly one of the grandest challenges of computational astrophysics.

3 Cosmic Jets

Let's take a closer look at the nearest rich cluster of galaxies called the Virgo cluster. Colorplate 10 is an enhanced optical image of the inner core of M87, which is the central elliptic galaxy in the Virgo cluster. This galaxy has about 1011 individual stars in it. You will notice a strange linear protuberance emitting an optical light coming out of the center of the galaxy. Another odd feature is that there are repeating bright spots along the linear structure and then it becomes wiggly at the end. The laws of stellar dynamics make it highly improbable that a subset of the stars in the spheroidal galaxy will decide to pile themselves up into this great pillar. And yet this M87 oddity has been known about since it was first noticed in 1918 by the astronomer Curtis. It was just stored away, as much of astronomy was until fairly recently, as one of those odd things that go on in the sky.

It took supercomputers and scientific visualization to help explain this sixty-year-old mystery. What is causing this phenomenon? The general theoretical framework for such phenomena was developed by Martin Rees and Roger Blandford, two brilliant astrophysicists in Cambridge, England, who in 1974 came to the conclusion that these linear systems must be a jet of gas being emitted from the center of the galaxy. This supersonic gaseous jet shoots outward, first through the gas between the stars then through the gas between the galaxies. By then similar jets had been observed in a growing number of other radiowave emitting galaxies as well.

Fortunately, there have been about one hundred years of experimentation and observation of supersonic jets of gas. The jet of gas in Colorplate 11 is a supersonic flow being emitted by the tailpipe of the Bell X-1 into the surrounding atmosphere. If you observe closely, you will see a series of bright spots similar to M87, which in this case are secondary combustion of leftover volatile products from the engine. The reason they are burning in the series of spots can be seen by going back to the laboratory work at turn of the century by Mach, who later became known for things like Mach waves and Mach numbers. Colorplate 12 shows a Schlieren photograph taken of gas being shot supersonically out of an orifice in a laboratory. In this closeup of the jet, one can easily see a series of "X" shockwaves that are induced into the flow. It is now understood that such standing crisscross shocks naturally should form in such gaseous flows as the nonlinear end-point of normal mode perturbations. The gas compression across these repeated "X" shockwave structures in the laboratory (on a scale of centimeters) is what is causing the secondary combustion in the Bell X-1 (on a scale of meters), and presumably what is causing the synchrotron electromagnetic emission in the jet in M87 (on the scale of light years).

The ability of physics to explain similar phenomena over scales of centimeters to 1018 centimeters is why it is such a powerful science. This scaling

property also means that one supercomputer simulation of such a gas flow can be simultaneously useful to laboratory researchers, aerospace engineers, and theoretical astrophysicists. Karl-Heinz Winkler, Michael Norman and myself decided to see if we could recapture this "X" phenomena by simply simulating the laws of gas dynamics on a supercomputer. This work was done in Munich in the early 1980s on the first Cray Research supercomputer in continental Europe.

What we discovered was that supersonic jets naturally form these internal regularly spaced shocks, which lead to the higher emissivity as seen in M87. Colorplate 13 shows the gas density as computed recently by David Clarke and Michael Norman for a fully three-dimensional magnetohydrodynamical jet. One sees both the series of bright knots in a linear structure as well as the beginnings of a wiggle at the end of the jet, remarkably similar to the optical image of M87 in Colorplate 10.

Furthermore, Norman, Winkler and I discovered from our supercomputer simulation that axisymmetric jets which had a lower density gas propagating into a higher density gas naturally formed a cocoon of returning gas around themselves. Thus, one would expect to often find jets with great plumes at their heads, such as in Colorplate 14.

This was just what was being observed by the VLA, the largest radio telescope array in the world. The VLA has 27 ten-meter antennas arranged in a Y pattern, all of which record the radio signals hitting their dishes from slightly different positions. By using large amounts of digital processing, one can reconstruct an image which would result from using an effective dish the size of the diameter of that entire array. For the most demanding reconstructions together with the further image processing needed to bring out faint features in the image, one requires a significant amount of supercomputer time. Colorplate 15 shows such an example of a nearby radio jet source called Centaurus A. The data, collected by Jack Burns and David Clarke, was transformed into an image at NCSA. You will see the twin jets emerging from the core of the galaxy and ending with large plumes, much as in the supercomputer simulation.

However, you can see that the plume is not axisymmetric, but that it meanders around the center line of the jet. So clearly something three-dimensional is going on. Now, the easiest procedure to try to capture this behavior is to look at the two-dimensional slab symmetry which allows for this sort of wiggling. Colorplate 16 shows such a numerical experiment in which Michael Norman of NCSA put a very small oscillation, where the thin slab of gas enters the grid from the left, and which gets amplified until you can see that it begins to wiggle the large plume, very similar to what you see in Cen A. This is caused by a nonlinear Kelvin–Helmholtz instability—the same instability that generates water waves and that makes flags flap. This visualization was created in collaboration with Donna Cox, a professor of

Art who also works at NCSA.

The full three-dimensional simulations are just now being done on the Cray Research supercomputers at NCSA. Mike Norman and postdoc David Clarke have created both a hydrodynamics and a magnetohydrodynamics code to study this situation. A single evolution can take up to 200 single processor Cray supercomputer hours (requiring a wall clock month to perform). The difficulty then is how to visualize the data so one can examine what one has computed.

Colorplate 17 shows four moments in a three-dimensional evolution rendered using a 3D imaging package called Dicer for the Macintosh. Dicer is made by Spyglass, which is a spinoff company that improved on NCSA imaging software for the Macintoshes and created a suite of commercial products. Colorplate 18 shows another graphic approach (developed by Clint Potter and his cc-workers at NCSA) in which we use the Connection Machine as a personal graphics workstation. One takes the three-dimensional computed jet density array into the Connection Machine and then uses the CM to compute a hundred and twenty different angles of view through this cube. If you take the video out of the back of the CM frame buffer, you can create an animation of the 3D image. Using these simulations, one can compute what the supercomputer-generated jet would look like to a radio telescope and then try to explain the specific jets one observes. This not only elucidates features of the jets, but also allows us to use the jets as probes of their gaseous environment. Such computations will be occupying the most sophisticated supercomputers well into the 21st century.

4 Supertelescopes

Now let me go back again to the VLA. The original design goal of the Very Large Array (VLA) of a dynamic range of 100:1 has now been increased routinely to 2,000:1 through digital processing techniques, and a dynamic range of 100,000:1 is achieved for point sources. The Very Large Array (VLA) can therefore be seen as an evolving "supertelescope" which becomes more powerful with time, by virtue of its coupling to new generations of more capable digital computers and new computational algorithms.

The radio astronomy community has a two-pronged strategy for creating the software required for this evolution. There are 200 researcher sites that have AIPS, the Astronomical Imaging Processing Software developed by the National Radio Astronomy Observatory (NRAO), running on their local computing resources. For most observations made with the VLA, minisupercomputers or workstations have adequate computing power and memory to take the output of the 27 dishes and reconstruct the radio image. But for about 15% of the most demanding observations, supercomputers are re-

quired to act as the "computational lens" of the telescope. When it became clear that NRAO would not be able to obtain its own supercomputer facility, NCSA worked with Dick Crutcher and his colleagues in the University of Illinois Astronomy Department to develop a new generation of software which was optimized to vector multiprocessor supercomputers. This led to the WERONG and MIRIAD software packages that are in use today by the radio astronomy community for creating "supermaps." This is the software which was used to create the map of Cen A in Colorplate 15. However, there are other classes of supercomputing imaging problems which are even more complex.

To show how critical advances in supercomputing are to this field, I have taken an edited version of one such problem from the "White Paper of the Computing and Data Processing Panel" prepared for the Bahcall Committee report. The NRAO has recently completed installation of 327 MHz (90 cm) receivers on the VLA. NCSA is working with NRAO to develop software capable of handling this data. Unique science addressable with this new capability includes steep spectrum objects and objects of large diameter and low surface brightness. Due to the two-dimensional geometry of the VLA, the samples of the visibility function are made throughout a three-dimensional volume, and the conditions under which a two-dimensional Fourier transform can be used to recover the source brightness fail, requiring much more expensive solutions. The simplest of these is a three-dimensional Fourier transform, producing a three-dimensional image "volume" whose axes are in direction cosines, and within which the desired image is found on a sphere of unit radius. Processing of the image follows the same procedures normally done in two-dimensional processing—for example, deconvolution proceeds in an entirely analogous fashion using the three-dimensional image with a three-dimensional beam.

A typical large project at 327 MHz will use all four VLA configurations with perhaps 12 hours observing in each. The imaging needs are exceptionally large. The simple 3D transform requires making a "dirty" map and beam, each 4096 by 4096 by 64 pixels, which with 4 bytes per pixel requires 8.5 GBytes memory if they are made in the most efficient, straightforward way. Fortunately, another approach is more efficient in memory, although at the cost of I/O and CPU. The image can be built up through a series of a large number of subimages of limited depth. The memory requirement is relatively more modest, about 135 MBytes per facet.

Deconvolution of the image is the next concern. In the simple, single cube approach, the procedure is straightforward, although highly consumptive of memory. No recall of the data is required. In the polyhedron imaging approach, deconvolution can be accomplished with much reduced memory, again at the cost of much increased I/O. A rough estimate is that perhaps 10,000 Fourier transform subtractions of clean components from the data

are required, along with 10 re-imagings of the entire field. That is, more than 2500 FFTs, each 1024 times 1024 will be required. Typically, three loops of self-calibration and deconvolution are required before satisfactory convergence has been achieved. Thus, all the operations described in the previous paragraph must be multiplied by three.

A rough estimate of the time required can be made: Using the Cray Research Inc.'s CRAY-2 supercomputer performance with a more modest case as a benchmark, and multiplying up by the ratio of database sizes and number of fields to process, results in a rough estimate of 250 DAYS for full processing. The time required is dominated by the gridding: Each visibility point must be distributed over about 100 adjacent cells, and averaged with all other visibility values within this cells, resulting in a computation-limited problem. We are confident that usable shortcuts will be found, as detailed studies of solving this computing problem have barely begun. For example, we can probably use a much less expensive gridding algorithm, which could cut processing by a factor of five.

It seems clear that the optimal approach will eventually involve parallelism. We could imagine 16 parallel machines, each comparable to, or faster than the CRAY-2 supercomputer. The data could then be distributed to each of these machines, each of which would be responsible for one subfield. A central processor will be required to handle the component model subtractions—the model comes from all 16 parallel processors. Factoring in these expected savings, and imagining future, more powerful machines, predicts this particular problem to be soluble with a few hours of computing time.

We're going to get a different sort of three-dimensionality when we observe the interstellar medium. This is the birthplace of stars, the nursery of stars that occurs in regions like the Orion nebula. The preferred observational mode is to use many different frequencies in the millimeter region looking at exactly the same place in space. Each frequency probes different physics, revealing the collapsing core of the star forming region or the bipolar outflows which are also characteristic of young stars. Millimeter wavelength observatories in the 1990s will be capable of receiving data in a thousand different frequencies simultaneously taken in two spatial dimensions. That makes a datacube of two spatial dimensions times frequency, instead of three space dimensional. This mode is going to be more and more common, so how do we build a computational environment to be able to do this routinely?

NCSA is creating an experiment in developing this environment as a part of its participation in the gigabit test bed project. The five national testbeds were pulled together by Bob Kahn, of the Corporation for National Research Initiatives, with funding by the National Science Foundation, DARPA, and vendors. The testbed NCSA is a part of is called BLANCA with fiber between Berkeley, NCSA, Wisconsin and AT&T facilities in New Jersey. We

are collaborating with the Berkeley-Illinois-Maryland Array which is upgrading the Hat Creek facilities outside Berkeley to what will become the world's fastest millimeter array telescope. The data from the telescopes will travel over the testbed (initially at T3 or 45 megabit/s) to NCSA's supercomputers where the image reconstruction and image processing will be done. Because of both the speed of the computers and the speed of the communications line, this will enable real-time observing. That is, while the telescope is still pointed at the object, the observer will be able to reconstruct the three-dimensional data set at NCSA, view it back in Berkeley on a supergraphics workstation in real-time, and decide whether or not to take more data and whether or not the observed parameters are right. Once this operational model is proven on this "national breadboard," the compute power will probably be made local to the telescope, with observers using the National Research and Education Network to connect remotely.

What is emerging is a planetary wide observatory. Observers all over the world will be linked by the NREN and its international sister networks into all the major astronomical observatories (both those on and off the earth). These will be linked as well into digital libraries containing all past observations and all the astronomical literature. Again most of the observations will be computed locally, with the national supercomputer centers concentrating on the most compute intensive ones. It will be the 21st century before this dream fully becomes reality, but then a new world of coordinated multiwavelength real-time observation will become possible.

5 Supermassive Black Holes–The Central Engine

Let us turn our attention to our own galaxy. Some of you will have been fortunate enough to have seen the beautiful ghostly band of the Milky Way on a really dark night from the Sierras or some other dark location. The Milky Way is really the plane of our galaxy which we are seeing edge-on. There is a big bulge of stars in the constellation Sagittarius, which is the direction toward the center of our galaxy. You may also have noticed dark patches in the Milky Way which are very obscured by interstellar dust. This dust absorbs essentially all visible light in the plane of the galaxy coming from further away than about one percent of the way to the galactic center. Thus, the activity near the center of our galaxy is completely obscured in visible light.

Fortunately, the interstellar dust is of such a size that it does not block radio waves. Using the WERONG software discussed above, astronomers Fredrick Lo and Neil Killeen produced a supermap of our galactic center which is shown in Colorplate 19. It shows the synchrotron radio emission (produced by electrons spiraling around magnetic field lines) in the central

couple of light years of our galaxy. To put the scale in perspective, that distance is roughly the same distance as from the sun to its nearest neighbor star (about 3 light years) and yet the galactic center is 30,000 light years away from us. What one appears to see in this image are streams of gas which appear to be spiraling into the very center of the galaxy.

This reminds us of the much more energetic activity we saw associated with the centers of elliptical galaxies. In Colorplates 10 and 15, we saw gaseous jets shooting out of the center of such galaxies. In addition, if you look closely at each of these images you will see a point source of emission located at the center of the galaxy. The search for the "Central Engine" at the center of galaxies and the enormously bright and distant quasars has been one of the Holy Grails of the last three decades of astronomy.

A standard model for this central engine has emerged. The idea is that there is a supermassive black hole (mass ranging from one million to one billion times the mass of the sun) at the center of a galaxy with gas spiraling in and forming a magnetized accretion disk orbiting the black hole. By a complex process which is still being debated, some of the gas escapes being swallowed by the hole and forms two outgoing supersonic streams of gas moving away from the black hole. In some cases the jet shoots out to a distance one hundred billion times the diameter of the black hole! So when you see those gigantic gas plumes on the scales many times the size of the galaxy (Colorplate 15), they are direct signatures of gas that left from the near vicinity of the central supermassive black hole.

Jim Wilson, a computational astrophysics pioneer at Lawrence Livermore National Laboratory, first used supercomputers (a CDC 7600) in 1972 to compute the general relativistic fluid flow in the vicinity of a black hole. A decade later, John Hawley, then one of my graduate students and now a professor at the University of Virginia, followed up with higher resolution runs on a Cray 1 supercomputer at the Max Planck Institute for Physics and Astrophysics in Munich. We performed axisymmetric computations of what would happen to a gas cloud which spiraled into a black hole. Colorplate 20 shows the sequence of events as the free falling gas is brought to a halt by a standing shock wave which forms because of the deceleration caused by the centrifugal barrier near the hole. The shocked gas then fills a toroidal region orbiting the hole. Because the boundary conditions specify that gas continues to fall toward the hole, the torus eventually fills up and the overflowing gas begins to move outward as two jets. These supercomputer simulations in the early 1980s showed that the existing basic theoretical structure of a fat torus with outflowing jet was not highly artificial, but a natural consequence of the general relativistic gas dynamics equations.

However, it was also clear from our simulations that just gas dynamics alone was inadequate to produce the observed magnetized jets that shot out to such great distances. We are almost sure that, since the accretion

disk is highly magnetized, the magnetohydrodynamic torques on this gas are what brings it near to the hole. Then, through a complicated interaction of magnetic field and radiation pressure coupled to the hydrodynamics, these two streams shoot out.

In summary, the central engine is apparently a magnetized disk of gas around a supermassive black hole. The black hole in this case is providing a passive gravitational field well for the matter to fall down into, to convert its energy from potential energy to kinetic energy, magnetic energy, and then back into kinetic energy as it goes out in the jet. The detailed working out of this phenomenon will clearly require the next generation of supercomputers.

6 Dynamics of Black Holes and Gravitational Radiation

Finally, I want to cover the most fundamental frontier in computational astrophysics, which is the dynamics of space and time itself. The basic notion of Einstein's theory is that any concentration of mass (like the sun) or energy (like a strong gravitational wave) will warp the surrounding space and distort the rate at which time passes. The warp in space causes trajectories of particles (like that of the earth) to deviate from a straight line, which interpreted in Newtonian terms means a force is acting between the mass and the passing particle. This perceived force is what has traditional been called gravity. If the bending of space is slight (like in the solar system) then Einstein's theory just reduces to Newton's theory of gravity.

If the mass is much more compact than the Sun, one can form a neutron star. This has the Sun's mass, but instead of having a radius of seven hundred thousand kilometers, it has a radius of about seven kilometers. The warping of space produced by neutron stars is still small enough that Newton's law of gravity is off by 10%. Thus, some investigators have been able to develop codes which solve the Newtonian hydrodynamics of these compact objects. While not accurate enough to give details of relativistic gravity, these computations do give a good overview of what we can expect matter to do in these extreme situations.

One such problem is the head-on collision of two neutron stars. Jim Wilson and I did the first computations of this problem with a fully general relativistic code in 1978. However, there was so much overhead in computing the general relativity, that we weren't able to study the hydrodynamics very well. Charles Evans, who's now at the University of North Carolina at Chapel Hill, created a very nice simulation of the Newtonian gravity collision while he was a postdoc of mine in 1986. The 3D visualization of tracer particles, shown in Colorplate 21, was created by Evans, Ray Idaszak, and Carl Hoyer. What happens first is that the neutron stars gravitationally fall

toward each other. When they collide, a great shock occurs which heats the neutron star material into a big mush. The outgoing shock front creates a lot of internal vortices and some very beautiful hydrodynamics. We can estimate the amount of gravitational radiation that such a time changing mass quadruple moment would emit if we were computing the full Einstein equations. Since this work, some investigators have begun to compute the complexities of a three-dimensional spiral-in collision. The real challenge is how to redo all this in the true curved spacetime which the mass creates.

To make progress in curved spacetime computations, we take the opposite extreme of clearing out all matter and just solving the Einstein equations themselves. If the warp becomes very large then some rather strange things occur. For instance, Einstein and Rosen showed in 1935 that one could have a hole in space which connected into another universe. This topological bridge connecting the two universes serves to act as a very compact mass. Since 1968, we have called it a black hole. The internal structure of a black hole cannot affect (or is casually disconnected from) the external world. From the outside you can't tell whether inside the hole is a massive object or in fact it has another universe hooked onto it. If you get too close to the surface of the hole (really just a featureless boundary in space), you will be pulled inside in an irresistible fashion. Light trying to pull away from the hole is retarded and redshifted and indeed pulled back into the hole if it is emitted too close to the black hole surface.

My Ph.D. thesis was on how to use supercomputers to solve the Einstein equations for black hole dynamics. The brilliant idea that this should be possible was first stated over thirty years ago in 1957 (!) by Professor Bryce DeWitt, now of the University of Texas, who was later to be my thesis advisor. The first efforts to carry out this program were made in the early 1960s on an IBM 7090, but it wasn't until 1973 that the single black hole spacetime was computed by DeWitt and his students (including me) on a CDC 6600. This was followed up by my work with Kenneth Eppley in 1978 on the collisions of two black holes using CDC 7600s. Today this work is entering its fourth decade with the NCSA computational relativity group consisting of David Hobill, Ed Seidel, David Bernstein, John Townes, and myself, working on the CRAY-2, CRAY Y-MP, and Connection Machine supercomputers.

Colorplate 22 shows a diagram of this Einstein–Rosen bridge, with the distance on the diagram correctly portraying the distances in space. This scientific visualization was created by Hobill, Bernstein, and myself with the help of Donna Cox and Ray Idazsak. The curvature of the diagram shows the actual warping of space as computed from the Einstein equations. The white line shows where the surface of the black hole is. The color of the surface represents the local clock rate, with the red showing that clocks tick slower in a strong gravitational field, whereas the blue shows where clocks

tick at our rate. Thus, this one scientific visualization shows precisely the spatial curvature and the time dilation caused by a black hole.

The reason to compute this simple spherical black hole is to create a reference spacetime, that is, one which is also known analytically. There are many issues of code calibration which this reference solution can help us with. The most fundamental one is how quickly the finite differenced solution approaches the analytic solution as the radial grid spacing is made smaller and smaller.

This is quite interesting even in the spherically symmetric case, since for technical reasons we choose to solve for the spherical black hole spacetime in a spacetime coordinate system which is dynamic even though the black hole gravitational field outside the hole is static. In Colorplate 22 you see this evolution, with the two white lines representing the surface of the black hole in the two universes. At late times the central region inside the hole stretches into a hypercylinder of constant diameter which is another test against analytically known behavior. It is fairly straightforward to get agreement between the supercomputer code and the analytic solution to one part in ten thousand.

However, the spherical black hole calibration code is just the warm up for the real current challenge, which is to relax one degree of symmetry and create a code for solving axisymmetric dynamic spacetimes containing distorted black holes and gravitational radiation. A certain amount is known theoretically about how such spacetimes behave in general, but there is no analytic approach which can generate actual solutions to the Einstein equations for these cases except in linearized situations. For instance, we know that a nonrotating black hole which is given a small distortion will gradually return to spherical symmetry by emitting gravitational waves of a characteristic form termed quasi-normal modes. These modes are of known wavelength and damping rate. However, this scenario has only been computed for linearized perturbations away from spherical symmetry. What happens if the hole is given a nonlinear distortion is currently unknown.

This brings up the need for computational symbolic manipulation. In general relativity you can't live without symbolic manipulation. Einstein's equations are so complicated that computer algebra is required to write them down explicitly in all but the simplest situations. More recently, symbolic computation has been used to generate not only the differential equations, but the numerical code as well. Not only does this represent a remarkable speed up in equation derivation, but reduces the number of mathematical and typographical errors during the code development stage. Our group has most symbolic packages available to the national community on a network server. The NCSA has also supported development of new capabilities for symbolic computing. For instance, while Steve Christensen was at the NCSA he collaborated with Leonard Parker of the University of Wisconsin

at Milwaukee on developing a new package for tensor manipulation called Math Tensor which sits on top of Mathematica from Wolfram Research, Inc.

The next stage after getting the differenced equations in the code correctly is to once again go through the several man-year effort of calibrating and refining the differencing methods until the code is accurate enough to do science with. This is carried out in two phases. First the nonradiative components of the gravitational field are compared with the spherical code described above. For instance, you distort the black hole and then see if at late times the evolution looks like that shown in Colorplate 22, so you know that the hole is settling down to a spherical final state. Second, you compare the radiative components of the gravitational field (which do not exist in the spherical case) with the results of the linearized perturbation theory for black holes, of which there is a large body of literature.

Now that the code is calibrated, that is, you know it calculates general relativity correctly, then what can you do with it? Well, wouldn't it be nice to come up to a black hole and kick it and see what it did? That is a little difficult to do experimentally! But a supercomputer and some scientific visualization makes it easy. In Colorplate 23, you see the evolution of placing a deformation in the curvature near the hole. You can see that the deformation breaks up into an ingoing and an outgoing gravitational wave train. Even at the speed of light, the outgoing waves have to pull out from the strong gravitational field and the loss of energy translates into a lengthening of their wavelength (the gravitational red shift). The ingoing train is swallowed by the hole and becomes static as the time becomes frozen. Again, the "Renaissance Team" of Cox, Idaszak, and our relativity group are responsible for the graphics.

The next few years will see a complete working out of the axisymmetric black hole spacetimes. For the first time, we will have accurate dynamics of strong field general relativity and will be able to lift the veil of ignorance which has surrounded this fundamental physical interaction for 75 years. I think that many times people supporting supercomputers don't realize that supercomputers are the ONLY way to solve the basic nonlinear laws of Nature for situations of realistic complexity, John von Neumann's basic point of almost 50 years ago. Thus, they can become an indispensable tool of the most esoteric theoretical investigations as well as a tool for practical engineering.

The Grand Challenge for the twenty-first century in this field is to develop community codes which can solve for the general four-dimensional solutions of the full Einstein equations. That is, the axisymmetry condition imposed in our current work must be relaxed. There is a national coalition of the leading researchers in numerical relativity being formed, which is studying how to build a three-dimensional dynamics code to solve the full Einstein equations. The gold ring problem that they are trying to solve is to compute

the spacetime representing the spiral-in collision of two black holes. This is the final fate of any black hole binary system. By coupling in matter, one could also compute the three-dimensional analogue of Colorplate 21, namely the spiral collision of two neutron stars. The great hope is that they can get the wave form computed before the experimentalists get the gravitational wave detector built so they can predict the signal rather than verify it. Accurate calculations of this spacetime will require a teraflop massively parallel supercomputer.

7 The Next Century

What one can see from the above tour of a few of the areas in relativistic astrophysics being investigated with high performance computers is that the next decade is going to see a coming together of theory and experiment to produce what the Bahcall Committee report terms "A Decade of Discovery." On the simulation side, the 1970s and 1980s produced a capability to model the dynamics of two spatial dimensional systems to a high level of accuracy. These computer codes have been highly optimized to vector architecture supercomputers. The next decade will see a similar flowering of fully three spatial dimensional simulations utilizing parallel architectures in order to get adequate speed. Most of these codes as we enter the 1990s are using either single or a few vector processors and are thus sustaining speeds measured in hundreds of megaflops. Within a few years, the minimum acceptable speed for a frontier code will be measured in gigaflops. Besides speed, these computations require large memories. The unit of memory for supercomputers in the 1990s will be the gigabyte.

Much of the 1990s will be invested in creating an entirely new technology infrastructure around the new generation of supercomputers to store, visualize, and analyze four-dimensional spacetimes routinely. The end user will need a capability of running a number of networked heterogeneous computers on a single problem. Large capital investments will need to be made by high performance computing and communications centers to support new architecture supercomputers, high bandwidth wide and local area networks, large on-line and archival storage facilities and interactive visualization systems. Similar investments must be made to provide graphics workstations and minisupercomputers to researchers from remote centers. Large software development investments will have to be made in the areas of distributed computing, massively parallel algorithms, user interface designs, and new scientific visualization modalities.

Although we can anticipate many more people entering the emerging fields of computational science and engineering in the next decade, I believe that a lot of their time will be taken up with learning how to effectively work

in this new networked technology base and in discovering how to extract science out of multidimensional spacetimes.

Thus, the dawning of the new millennium should see the beginning of a golden age of computational science. The community will have absorbed this huge new dose of technology and have settled down into routine use of massively parallel systems operating in a gigabit/sec nationally networked environment. The information infrastructure will enable scientific research and collaboration to be performed in a manner almost unrecognizable today.

Acknowledgement

I would like to thank David Hobill, Radha Nandkumar, Mike Norman, and Janus Wehmer at NCSA, who were very helpful to me in putting this paper together.

References

[1] Bahcall Committee Report.

[2] Evans, C. R. "Gravitational Radiation from Collisions of Compact Stars," in 13th Texas Symposium on Relativistic Astrophysics, pp. 152–156, 1987, ed. M. P. Ulmer (World Scientific Publishing Co. Pte Ltd.: Singapore).

[3] Norman, M. L., Clarke, D. A., and Stone, J. M. "Computational Astrophysical Fluid Dynamics," to appear in Computers in Physics, spring '91.

[4] Smarr, L. "An Approach to Complexity: Numerical Computations," Science, 228, pp. 403–408 (1985).

[5] Smarr, L, and Hawley, J. "New Paradigms for Black Hole Accretion from High Resolution Supercomputer Experiments," In American Institute of Physics Conference Proceedings 144: Magnetospheric Phenomena in Astrophysics, 263-290, 1986, ed. R. I. Epstein and W. C. Feldman (American Institute of Physics: New York).

[6] Smarr, L., Norman, M., and Winkler, K.-H. "Shocks, Interfaces and Patterns in Supersonic Jets," in Fronts, Interfaces and Patterns, 1984, ed. A. R. Bishop, L. F. Campbell, and P. J. Channel (North Holland), and in Physica 12D, pp. 83–106, 1984.

[7] Wilson, J. R., Astrophys. J., 173, p. 431 (1972).