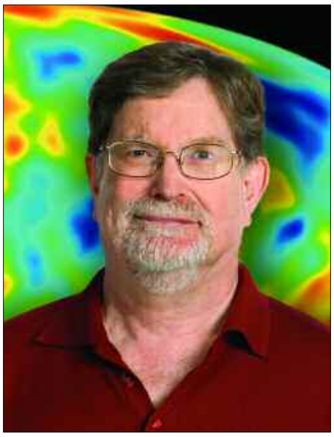
COSMOLOGY and COMPUTATION


LBNL

Dr. George Smoot recently took time from his overcrowded schedule to talk with LBNL's Jon Bashor about life since the Nobel Prize, the growing role of computation in cosmology, and what the future holds.

SciDAC Review: *First of all, congratulations again. Can you take a few minutes to talk about "life after the Nobel?" What has been the biggest surprise?*

Dr. Smoot: The biggest surprise was winning the Prize, but the second biggest surprise was how intense the related activities turned out to be. There were eight days of activities, from morning to night, in Stockholm. A German TV crew was there and asked me about the difference between receiving an award in Germany and the Nobel Prize. In Germany, you have a half-hour meeting with the chancellor. In Sweden, there are receptions, concerts, a banquet, and parties. It's a very big deal and you get to meet royalty—crown princesses, the king and queen. It goes on and on.

Another interesting aspect is you suddenly become an expert on everything. Recently in Berkeley there was a symposium on energy sustainability, followed by a Nobel laureate symposium. Afterwards, we went outside to have our photo taken for *Vanity Fair.* And since I've been wearing a suit so often, I've gotten an interview request from *Esquire*. The pile of requests for interviews and appearances is now about six inches high.

The award is in recognition of results you first reported in 1992—results which you said at your press conference could be analyzed on one of today's laptops. Since the late 1990s, you and your colleagues have used significant resources at DOE's National Energy Research Scientific Computing (NERSC) Center, including running the first application to use all 6,000 processors on the IBM supercomputer. Could you talk about how high-performance computing has advanced your research?

Computing has played a big role from the very beginning of what is a long series of experiments, and the computing requirements have grown with each experiment. Some of the first experiments—the ones that discovered the dipole anisotropy—took digital data on board a U2 airplane. We then used LBNL facilities to read and process the data. We then used balloon-borne experiments—two flights each in the northern and southern hemispheres to get sky coverage. Again, this was digital data. On one of the flights, the gondola and data were lost in the Brazilian rainforest. The parachute got hung up on the jungle canopy so that the payload did not reach the ground, until it was found by a palm-heart poacher a year and a half later. The tape had mold and mildew on it, but LBNL computing staff were able to get the data from it.

Computing has played a big role from the very beginning of what is a long series of experiments, and the computing requirements have grown with each experiment.

For the Cosmic Background Explorer (COBE), our prototype system was an LSI computer with a TTY teletype modem and the software was based on that used for a previous experiment. Although the bulk of data processing was done at NASA's Goddard Space Flight Center, we developed many (about 40% or more) of the algorithms here and then shipped them there. The actual analysis of the processed data was performed primarily at two places—Goddard and LBNL. At LBNL, we had a farm of 10 DEC Alphas and the instrument expertise—me, some graduate students, and some post-docs. We also performed the role of remote location back-up of the data and were called upon on one occasion to send back a significant amount of data that was accidentally lost.

At the time, our software people predicted we could do the entire job on one of the DEC Alphas, but the task grew and the

In Brief: Dr. George Smoot and CMB Radiation

In the wee hours of October 3, 2006, physicist Dr.George Smoot of Lawrence Berkeley National Laboratory (LBNL) received a telephone call from Sweden notifying him that he and NASA's Dr. John Mather were co-recipients of the 2006 Nobel Prize in Physics "for their discovery of the blackbody form and anisotropy of the cosmic microwave background radiation." The phone call was both a well-deserved recognition of their work studying the evolution of the Universe and the start of a seemingly endless stream of invitations to give lectures, requests for interviews, and inquiries about participating in special events.

A graduate of Massachusetts Institute of Technology (MIT), Dr. Smoot came to LBNL in 1974 and is most famous for his research on the cosmic microwave background (CMB) radiation ("Mapping the Universe to the Beginning of Time," p14). This radiation is thought to be the relic of the intense heat of the early Big Bang. In addition to his research at LBNL, Smoot has been a professor at University of California—Berkeley since 1994.

In April 1992, Dr. Smoot made the announcement that the long-sought variations in the early Universe had been observed by the Cosmic Background Explorer (COBE) Differential Microwave Radiometers (DMR) experiment team he led. NASA's COBE satellite mapped the intensity of the radiation from the early Big Bang. The team was the first to find evidence of the tiny variations that are the seeds on which gravity worked to grow the galaxies, clusters of galaxies, and clusters of clusters that are observed in the Universe today.

As Dr. Smoot has explained, "**The tiny temperature variations we discovered are the imprints of tiny ripples in the fabric of space-time put there by the primeval explosion process. Over billions of years, the smaller of these ripples have grown into galaxies, clusters of galaxies, and the great voids in space.**"

Since then, subsequent cosmic microwave background observations, including data from dozens of ground-based and balloon-borne experiments, and from the Wilkinson Microwave Anisotropy Probe (WMAP) satellite, have confirmed and refined the original maps. Because of the massive amounts of data involved, the data from many of these experiments have been analyzed using supercomputers at the Department of Energy's National Energy Research Scientific Computing (NERSC) Center.

Over the past ten years, Dr. Smoot and his colleagues studying CMB data have used nearly five million processor-hours and tens of terabytes of disk space on supercomputers at the NERSC Center in Berkeley. As work progresses on the European Space Agency's Planck satellite to be launched in 2008, researchers in LBNL's Computational Research Division have developed new software to handle the massive amounts of data to be generated. Named for Max Planck, the father of quantum theory, the satellite is designed to map CMB temperature and polarization fluctuations with unprecedented resolution and sensitivity, but the enormous volume of data this will generate poses a major computational challenge.

After COBE, the goal was first to capture the finer details of the temperature fluctuations, requiring experiments with many times the resolution of COBE, and this was realized by subsequent experiments. After the recent ultra-high-precision WMAP results, the next goal, targeted by the upcoming Planck satellite, is to detect the even fainter polarization fluctuations, which requires experiments with 100,000 times the number of observations. But analyzing this amount of data using the same algorithms as were applied to the COBE data would require millions of processor-years.

The angular resolution of sky maps is measured in arcminutes, with one arcminute equaling 1/60 of a degree. The width of the Sun or the full Moon viewed from Earth is about half a degree. While the original COBE maps used by Dr. Smoot's team had an angular resolution of seven degrees, maps generated from Planck data are expected to boast an angular resolution of five to ten arcminutes and a sensitivity of one millionth of a degree.

This data explosion has spawned a computational challenge, which is driving analysts to high-performance computing centers like NERSC, which now supports around 100 analysts from a dozen experiments.

dataset expanded. The data unfolded to 10 times the size of the dataset we took with the experiment. It turned out to be a much bigger job and, in fact, we did a more sophisticated analysis.

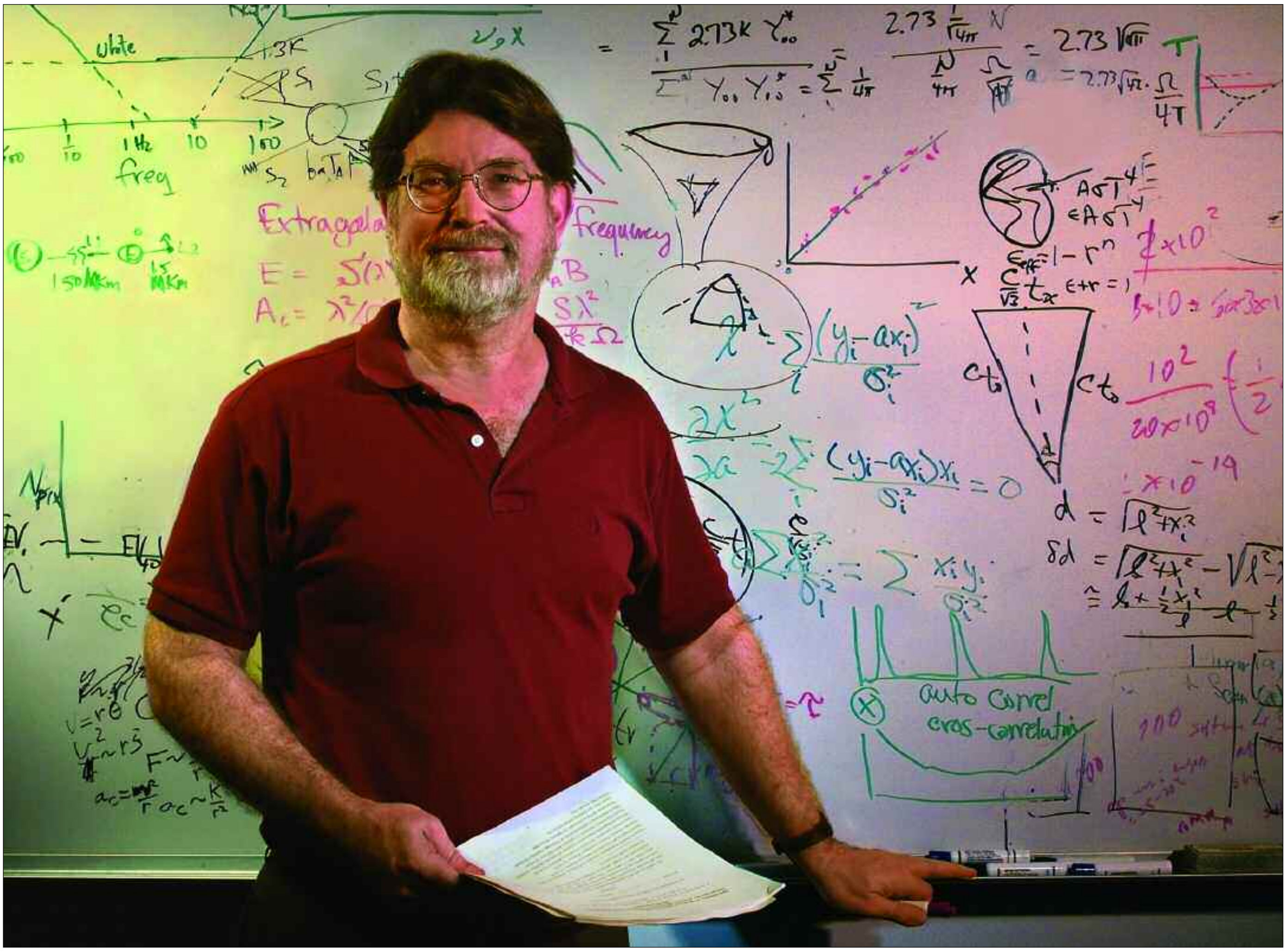
With COBE, we had a map with 6,144 pixels. With the Wilkinson Microwave Anisotropy Probe (WMAP), the map had three million pixels and Planck will give us a map with 50 million pixels. Because of the amount of data, we are now forced to do approximations in Monte Carlo. Using exact methods, it would probably take longer to analyze all of the data than it took for the Universe to develop.

As we increase the precision of our approximations from 50%, to 10%, to 5%, and now to 1% percent, the iterations go on longer. It's a major computing problem as to how we compress the data to reproduce a simulated universe that is indistinguishable from our Universe. Instead of solving this problem with brute force, we have to be clever, and that means clever algorithms and implementations. Our goal is to get it down to a small number of approximations that are as close to the real Universe as we can get.

So how close can we get? When I was a student, they used to have "mixers" to bring together boys and girls from different schools, like Harvard and MIT. We'd start out with all the boys on one side of the room and all the girls on the other. To bring us together, the organizers decided that each group should move half the distance across the room, and keep doing that until we met. Well, the mathematicians right away said that we'd never get there, but the engineers from MIT said "We'll get close enough."

Looking five years down the road, how do you see the state of computational astrophysics and cosmology?

We are looking five years down the road with Planck, which is launching in 2008. Two years of data will take us to five years down the road while we are processing it. We know it will be a tremendous computing challenge. Then there are other issues.



R. KALTSCHMIDT, LBNL

Figure 1. *In addition to his astrophysics research at LBNL, Dr. Smoot serves as a professor of physics at the University of California–Berkeley.*

There are a lot of interesting problems in cosmology that have to do with large structures that require simulations that strain you in terms of pure numbers and range. If you start with a universe, even though it's really compressed and the materials are uniform, if you trace it over time, there is density contrast that's huge. You will take what's uniformly spread out and crunch it down by very large factors. So you have this range.

For five years, or probably 10 or 15 years, cosmology is going to stress large-scale computing in a serious way. If you look at a web of the Universe, you have a large number of particles, 1,012, a tremendous range. We might be able to solve up to 108 or 109, but you usually run out of computing power.

For a while people tried to build machines that will solve gravity and stability problems. Every object has an effect on every other object. But it's the nearest one that affects you the most. There are specially designed computers to attack this problem. But the fact is that general-purpose computers were getting their flops faster than people could build special systems. To use these general-purpose computers, you have to come up with clever algorithms. You do all these clever things. If you look at these simulations showing the cosmic web, you see this thing where it's pretty uniform, a spider web, and the cosmic web is made up of clusters of galaxies.

You can run for a while and see the large-scale structure. Then you need to follow one cluster and see how the Galaxy is formed. But there are too many steps to run and people use various techniques to make approximations. But that's not all that you want to do. Let's say you want to trace light rays from the beginning of the Universe. You have to start back from 13 billion years ago, then four billion years ago, a billion years ago. It's too complicated to solve the entire problem, so we approximate by taking slices from different time steps. Certain kinds of problems are best solved by laying out a set of general rules and solving the problem locally. Right now we do an approximation because it's too complicated a problem to be solved in its entirety.

Two years ago, to simulate processing an entire year's worth of Planck data at the single most cosmic microwave background (CMB)-sensitive frequency, our team used 6,000 processors on NERSC's IBM supercomputer for nearly two hours—the first time virtually all of the processors were used on a single code, mapping 75 billion observations to 150 million pixels. They had to dedicate Seaborg to get just this one wedge because you want to go from early Universe to present, you want to be able to trace the line down. It took a huge amount of effort and time.

You asked me what I see in five years. In five years, I see astrophysics challenging computation in a serious way. More than five years. It's clear that it's a humongous challenge already and will only get larger as experiments beyond Planck will generate datasets that are orders of magnitude larger.

We've heard that you plan to donate a significant portion of the prize money to establish a new center at LBNL and University of California–Berkeley. One component of this effort would be to bring together scientists from the experimental and computational sides. Can you elaborate on this?

We've reached a point in cosmology—with the CMB discovery, the supernova group, large-scale structures, surveys, and simulations—where we know a lot about cosmology and have brought the parameters down to around the 2% level. We can expect in the next few years to get down to the 1% level. When you get down to the 1% level, you are actually testing your knowledge well.

Most things in your life are alright with a 1% parameter and you are happy with it. For example, your pant size. If it's wrong by 10%, they will fall down or you can't button them. If it's 1%, it can be a little tighter or a little loose, but it doesn't matter. You can use a belt. There are lots of things where if you get to 1% then it's good enough. You are reaching a threshold where you understand things reasonably well. When you look at things carefully at 1% level, you can tell whether your model is right or wrong. The same is true for the science about the Universe. When we get to the 1% level, we will know whether we are right about what the Universe is. You have something called the standard model, and then you go down and say what physics you have to assume. You have to assume inflation, you have to assume dark energy, and I made this list of eight things.

Understanding the simplest form is like understanding the ideal gas law. We know what makes it work—I plug in numbers and run them and we can compare them with the observations. When we get to 1%, we will test things like we do using the ideal gas law. We are in a situation where testing will tell us if general relativity is wrong or right. Is there dark matter or something else weird going on? Does the dark matter cause galaxy structure to be different? There are a lot of questions we can answer once we get to that 1% precision, provided that we can do the computation. We have to do the simulations that can reach that 1% accuracy.

I thought this new center is a tremendous opportunity to define what those questions are and to see what the modes of attack are. We also want to have a lot of education and outreach.

For five years, or probably 10 or 15 years, cosmology is going to stress large-scale computing in a serious way.

The public finds this interesting. People who become engineers and computational scientists start with a problem they are excited about. If you present it to them in a way through posters and websites, then you can get them interested. I got an email from a student who failed my course the first time he took it. He got a D. He was allowed to retake it and he did. Now he's going to be a plasma physicist because he got motivated, so he was able to understand how to do it. Before, he thought science was hard. You know what they say: physics saves lives, because they make pre-meds take physics to keep certain people out of medical schools.

Figure 2. *Nobel Prize winner Dr. Smoot visited Cannes to view the mirrors of the European Space Agency's (ESA) Planck satellite. Planck will serve in further studies of CMB, the relic radiation from the Big Bang.*

The Universe is predicted by simple rules and you can create a model. Then you can go forward in steps. This is where computer visualizations allow us to create pictures and movies and really demonstrate graphically what happens. That attracts people to the technical field. This is a field where you can apply mathematics and reasoning. It's not only the intellectual challenge of understanding the Universe but also the challenge of getting people interested in the issue.

In newspaper stories that have appeared since the Nobel Prize announcement, there has been a lot of mention about how your family moved around, from Alabama to Alaska, due to your dad's job with the U.S. Geological Survey. How did you end up at LBNL?

There were tremendous opportunities here, and a can-do attitude. Not only was I given the tools and resources to do research, but I was also shown a style for doing research that is very different. I came here from MIT and we had a way of doing things at MIT that was very impressive. But here it was a very different style that came from our founder—Ernest Lawrence—and Robert Oppenheimer. It was: pick out the best science you can do, and do it. That was so liberating. Do the best science you can. That was what helped me think about science that was out of the ordinary; cosmology wasn't really an established field back then. It's no wonder the Free Speech Movement started here. It's the "Free Science Movement."

Finally, if and when your life settles back down, what's next on your horizon?

I've heard from most people it doesn't.

Thank you for taking the time to answer our questions. ●

Further Reading

"Mapping the Universe to the Beginning of Time," p14