Proposal for an earth science computing project at Stanford

Rick Ottolini

ABSTRACT

An earth science computer project would pull together disorganized computer research in various labs at Stanford. The essential part of the project is a faculty-level research scientist able to conduct cutting edge research both in earth science and computing. Research areas include data processing, simulation, visualization and basic computer science.

FORWARD

I delivered this proposal to SEP sponsors at the Annual Sponsors Meeting in May 1988 and made a formal presentation to the Stanford School of Earth Sciences[1] a month later. My intent for the Stanford audience was to propose a solution to what I and several faculty perceive to be a growing problem at Stanford—the lack of direction in earth science computing. In addition, I had a personal reason for designing a job promotion for myself. I made the presentation to SEP sponsors to scout out moral and financial support. Since many of these ideas grew out of SEP-related research, I thought the proposal would be of interest to SEP sponsors. To my surprise the proposal drew more interest outside Stanford than within. Several sponsors and other institutions said they were thinking along similar lines and asked for a written copy of the proposal. Here it is.

[1]The four earth science departments at Stanford— Geology, Geophysics, Applied Earth Sciences, and Petroleum Engineering— are joined into their own School.

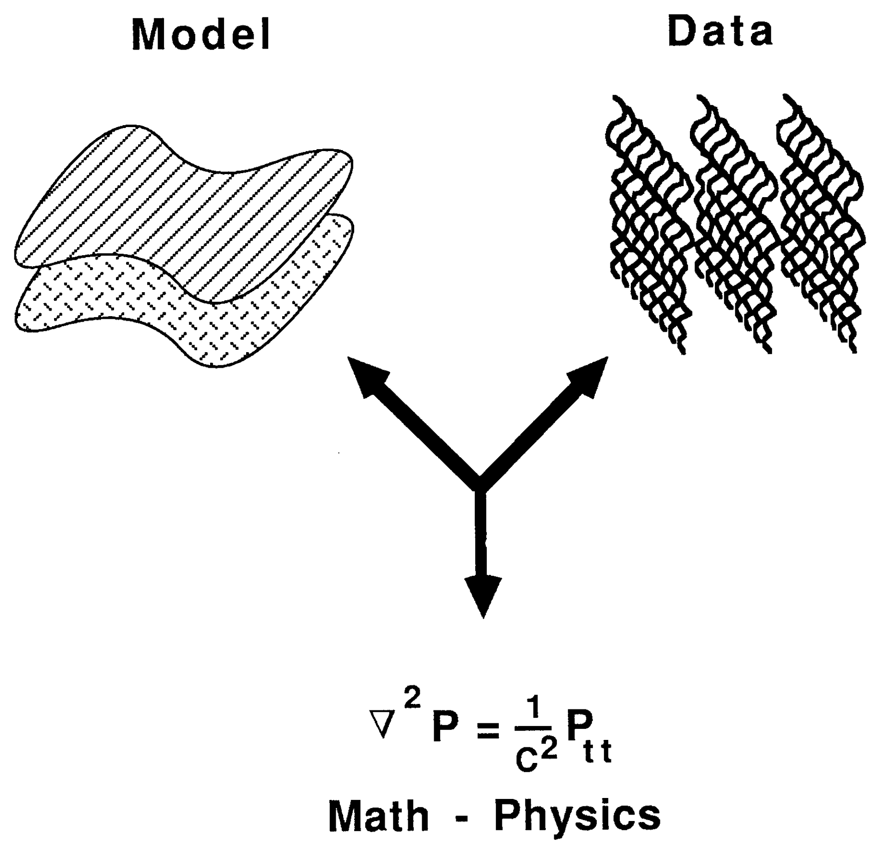

FIG. 1. Scientific computing in the twenty-first century will be *MacDraw*-like. Images of the data, geological and mathematical models interlock. Adjusting one instantly updates the other.

COMPUTER EARTH SCIENCE IN THE TWENTY-FIRST CENTURY

In the twenty-first century, the earth scientist computes by manipulating images on a console screen. The images are data, models of the data, or models generated from first principles of physics, mathematics and geology.

The images *move*— change content, shape, color. A different part of the data is displayed. A model parameter is changed. A time history is shown.

The earth scientist has complete and exquisite control over the content of the images. Yet there is no programming in the conventional twentieth century sense. The earth scientist draws an image, types an equation or adjusts a flow-graph. The screen image is intimately linked with the underlying numerical and geological model. Changing one rapidly and accurately changes the other.

The parts are here: interactive graphics, physics and mathematics of geology and data, and the supercomputing horsepower to drive it all. What is absent is someone to put it all together and make it conveniently available to everyone. The twenty-first century is only twelve years away.

HIRE A RESEARCH COMPUTER SCIENTIST

A computer science staff position should be created in the Stanford School of Earth Sciences. The position is responsible for the acquisition, development, and dissemination of new computer technology, particularly software.

This position is an important component of School plans for the twenty-first century. State-of-the-art computing is necessary for research and education in the earth sciences. The creation of a high level position in the School would give Stanford prominence in earth science computing.

This position builds upon existing strengths in the School. The School is located in an university and industrial region that are at the forefront of computer science. Some of the School's research consortia already use advanced computer technology. The School already teaches some earth science computing courses in UNIX, geomathematics, and geostatistics. The School already has a computer manager to administer computing activities.

This position solves problems in the School. Many professors lack access to computer expertise to solve their research problems. Many labs share computing problems and would benefit from a coordinated solution. An accumulating, long term research program is needed in contrast to students who complete a project and then leave.

Why a research level position? Can't a computer technician do as well? No, by definition, a PhD-level of expertise means someone capable of recognizing *significant* problems and has the skills to solve them. A person skilled only in computer science may not understand aspects of earth science and vice versa.

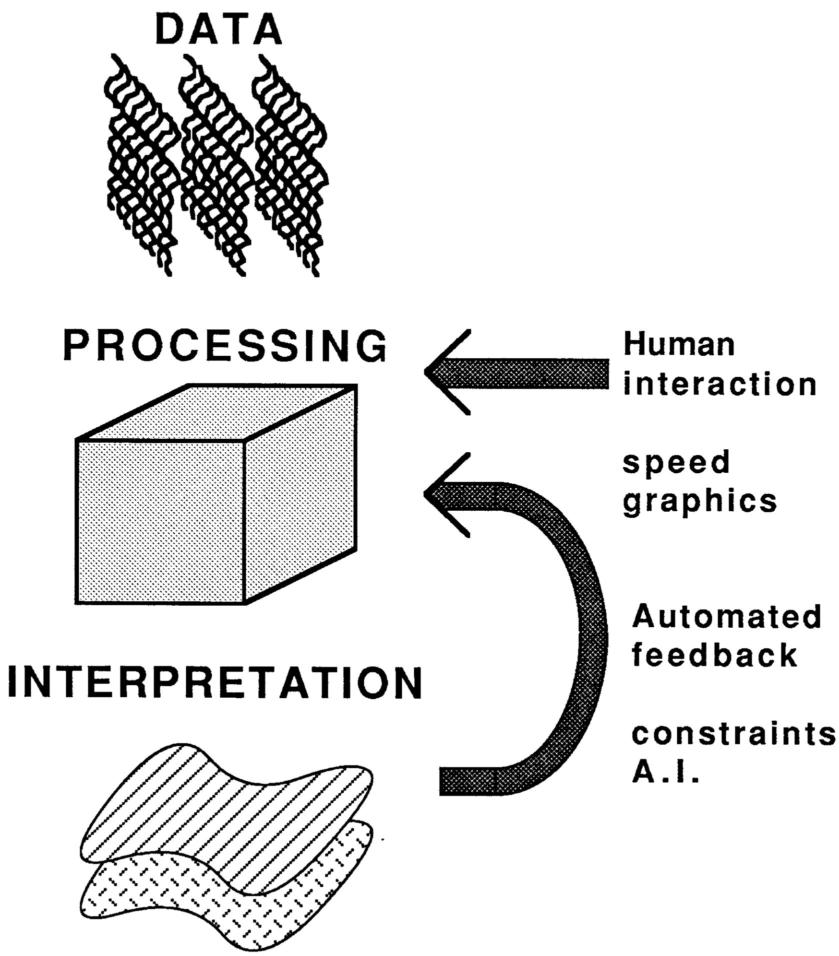

FIG. 2. The application of the interpretation tools of interactive graphics and pattern recognition to data processing is called *interpretative processing.* Interpretative processing aims toward the contrasting goals of automating as much of the processing sequence as possible, yet allowing human interaction at all level.

There is a conflict between devoting resources to writing software in house and obtaining it from elsewhere. By the time technology is available from a vendor it is often no longer at the cutting edge of research.

Duties

The primary duty is research in earth science computing. Research generates new scientific results and solutions to computing problems. Research areas are covered in the next section of this proposal.

The second duty is education. Stanford's most important product is trained earth science professionals. The person would supervise theses and teach courses. A side product of computing research could be educational software.

The third duty is communication. Research results are presented in seminars and papers. The person interacts with professors, the School computer committee and university computer services.

The fourth duty is to acquire new computing resources. The person keeps abreast of latest products, solicits donations from industry or buys when necessary.

The last duty is to raise funds. These funds supplement the position's salary, pay assistants, fund grad students and buy new equipment. Fund raising sources are discussed later in this proposal.

RESEARCH AREAS

Earth science computing research includes data processing, simulation, visualization and basic computer science.

Data processing is the conversion of raw data into a geologic model. Large datasets and accurate calculations require supercomputers. Interactive data processing gives better control over the data and requires computing resources.

Simulation gives insight into data and geologic processes. Seismic data, petroleum reservoirs, sedimentary basins and tectonics may be modeled. Simulations are increasingly multi-dimensional and dynamic.

Graphical display is essential for grasping large datasets and models. It is useful for understanding spatial and temporal relationships. Display is a method of quality control. Inconsistencies and gaps stand out.

Basic computer science supports the above applications. New hardware and software are continually invented and must be mastered. Graphics is the foundation of modern computing. Numerical algorithms bridge science and computation. Non-numerical algorithms, such as symbolic computing and databases play increasing roles in solving problems. Software engineering integrates these and makes it available to everyone.

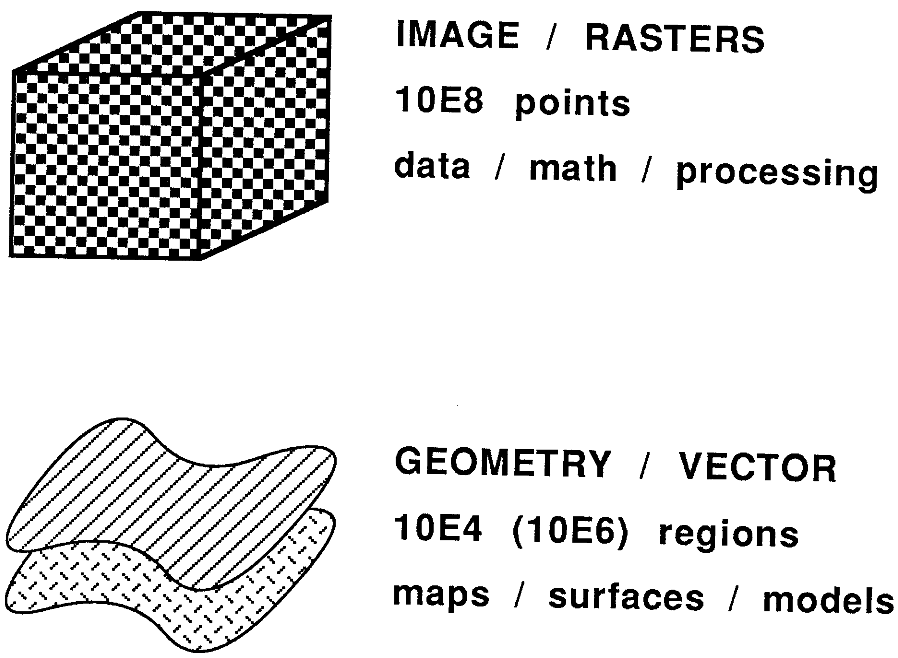

FIG. 3. There two major visualization techniques, each with its characteristic data, hardware and software. Raster hardware displays regularly sampled data. Vector hardware displays irregular regions.

Funding

The initial budget is for salary and computing resources. Initially the position should be funded from School laboratories that would make immediate use of its results such as the Stanford Exploration Project (SEP), the Stanford Reservoir Characterization Project (SCRF), and the Sedimentary Basin Simulation Project (SEDSIM). In time, the position would generate external funding and grow to include assistants and its own industrial consortium.

The position should have the authority to pursue external funding for salaries and computing resources. Sources include:

- The earth science industry, particularly oil companies.
- The computer science industry. Donations are encouraged in exchange for a petroleum industry audience and demonstration software.
- The National Science Foundation funds a supercomputing visualization initiative. The Defense Advance Research Projects Agency and NASA have similar projects.
- Government agencies such as the National Science Foundation, US Geological Survey, and Integrated Research Institutions in Seismology.
- Software sales. Industry, government, and education are clamoring for good software.

There is the danger of promising too much to a commercial funding source. Then the position may turn into a consulting service to sponsors and neglect research.

Qualifications

The person should be able to conduct publishable research in computer science and one of the earth sciences. This implies a graduate degree in computer science or a graduate degree in earth science with a strong computer background.

Candidates may be sought from academia, industry or government in either the computer or earth sciences. It is difficult to attract someone whose initial background is computer science due to severe competition in that discipline. Oil companies tend to be poor innovators in computer science, though there are exceptions. Government agencies with strong earth science computing such as JPL and NASA should not be overlooked.

The candidate would propose and begin work on a couple of well-focused projects which coincide with needs in one or more of the School's labs.

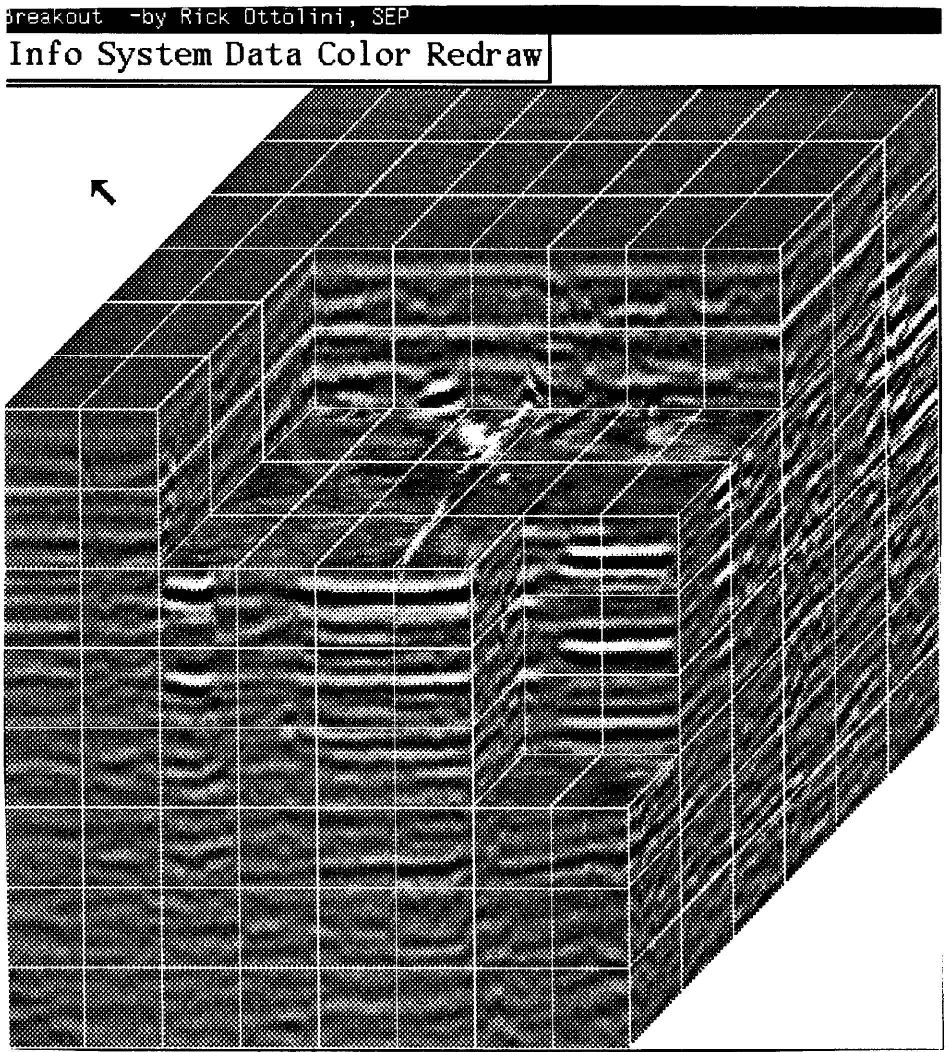

FIG. 4. One method of displaying the interior of data volumes is by *sculpting.* This is a three-dimensional seismic image of marine sediments. One carves the image by adding or deleting sub-blocks.

MODELS AT OTHER UNIVERSITIES

Stanford CASIS

The Center for Aeronautical Sciences and Information Sciences consortium (CASIS) in Stanford's Electrical Engineering Department applies computer science to the space sciences (including space geophysics). Projects include hardware and software systems, graphics, artificial intelligence and numerical simulation. They have a research staff position, grad students, computer resources and an annual meeting. Sponsors include NASA, aeronautical and computer companies. I consider this a successful model.

University of Houston

The University of Houston Seismic Acoustics Lab has industrial consortia in super-computing, graphics, and reflection seismology. Each has research staff, grad students, computers and annual meetings. Unfortunately, they expend much of their resources in supplying services to financial sponsors such as supporting computer codes. I consider this undesirable if it can be avoided.

Nancy, France

Professor Mallet, at the Geologie Ecole in Nancy, France, runs an earth science computing center with research into geostatistics, expert systems, three dimensional geological modeling, etc. Most of his financial support has been from the French government and the Common Market, though industrial contact is increasing.

ADDENDUM: SOME RESEARCH PROJECT SUGGESTIONS

The possibilities are large. These are some items I would work on if in this position. Though these have roots in geophysics, they are applicable to geology and petroleum engineering research needs.

Image browser

An image browser interactively displays a pre-computed set of images. There is interactive control of the display in order to

(1) see a different view of the image,

(2) enhance a feature by adjusting color, contrast or size,

(3) juxtapose separated portions of image for comparison,

(4) select and mark portions of the image, and

(4) simulate 3-D and animation.

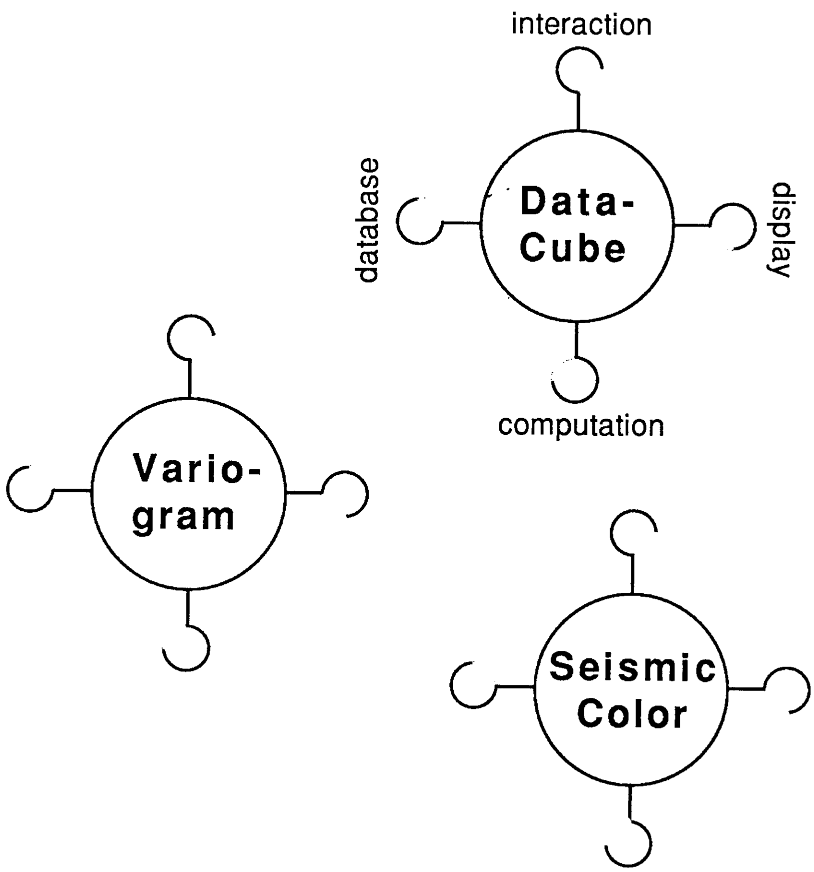

FIG. 5. The *geotoolkit* contains reusable program building blocks. Blocks are *data abstractions.* Parts describe object properties, how it is displayed and how it computes. The toolkit will be a geologic layer atop existing toolkits.

There are two major graphics representations of images. The first is regularly sampled arrays of numbers. Geophysical surveys, photo-geology, and simulations fall into this category. Raster graphics is used to represent these. The second are less regular datasets such as lines, surfaces, volumes and combinations thereof. Geologic models and mathematical functions fall into this category. Vector graphics is used to represent these.

Graphics hardware tends to be optimized about one or the other representation. Fast raster hardware is called an image engine, e.g. *Pixar*. Fast vector hardware is called a geometry engine, e.g. *Silicon Graphics*.

This dichotomy occurs in software too. Raster software is reasonably well developed in the SEP and Remote Sensing Labs (StanSearch) Vector software is relatively weak. Therefore, image browser development should strengthen vector applications and attempt an unification of these two approaches.

A properly designed browser is general enough to work with most datasets, yet contains functions specialized for earth science datasets. The SEP movie program is an example. Although designed for seismic images, it works reasonable well with any volume of numbers.

A significant research problem is the visualization of volume interiors. Do you carve away material or introduce transparency? The answer is partly dependent on the spatial structure of the dataset.

The geotoolkit

Toolkits are building blocks for easier program writing. They include data abstractions and code libraries. Toolkits already exist for numerics and low-level graphics. Toolkits are needed for geological entities, their graphical representation and numerical manipulation.

Consider a reflection seismic data toolkit for example. Some parts would be data oriented, e.g. seismic-trace, some would have a more graphic flavor, e.g. seismic-color-spectrum, and while others would have a numeric flavor, e.g. trace-gain.

Tectonics simulation

The 3-D animated simulation of global and local tectonics is an important research and education tool. This could be a geologic layer atop existing engineering software for volume modeling and deformation.

Tectonics simulation is *constraint-based.* The pieces must be manipulated in space and time without gaps or overlaps. Sometimes such inconsistencies are geological significant, e.g. pull-apart-basins.

AFTERWARD

This proposal received a mixed reception at Stanford and has lost momentum. Some of the faculty see little need for computers at all in the geological science, except as word processors. The balkanization of Stanford earth sciences into many departments, laboratories, and buildings makes it difficult to put multi-disciplinary projects together. I expect the computer software problems in the School to grow as more labs acquire powerful hardware. Stanford has a chance to become a leader in earth science computing, if it will grasp the opportunity.