



COMPUTING:

Screen Savers of the World Unite!

Michael Shirts and Vijay S. Pande*

Recently, a new computing paradigm has emerged: a worldwide distributed computing environment consisting of thousands or even millions of heterogeneous processors, frequently volunteered by private citizens across the globe (1). This large number of processors dwarfs even the largest modern supercomputers. In addition to the scientific possibilities suggested by such enormous computing resources, the involvement of hundreds of thousands of nonscientists in research opens the door to new means of science education and outreach, in which the public becomes an active participant.

A handful of projects have already demonstrated how such large-scale distributed computing power can be utilized. For example, SETI@home has totaled over 400,000 years of single-processor CPU time in about 3 years in its search for alien life (2). Similarly, distributed.net has used the power of this huge computational resource for the brute-force cracking of DES-56 cryptography codes.

Virtually any other computationally intensive project could be aided by distributed computing, from the simulation of nuclear reactions or star clusters to atomic-scale modeling in material science. Perhaps the most exciting possibility, however, is in the biological realm. In the last few years, the huge amount of raw scientific data generated by molecular biology, structural biology, and genomics has outstripped the analytical capabilities of modern computers. Novel methods, algorithms, and computational resources are needed to process this wealth of raw information. For example, we need to compute the structure, thermodynamics, dynamics, and folding of protein molecules, the binding ability of drugs, and the causal events in biochemical pathways. Many of the newest distributed applications have thus focused on biological systems.

Both SETI@home and distributed.net tackle so-called "embarrassingly parallel" problems, in which the desired calculation can easily be divided between many computers. For example, SETI@home looks for alien life by Fourier-transforming radio telescope data from different parts of the sky. These chunks can easily be assigned to different computers to be processed. However, not all problems are so easily broken down into independent parts ("parallelized"). Just as having 1000 assistants does not necessarily mean that one's work will be done 1000 times faster, the great challenge for distributed computing is the development of novel algorithms that allow calculations previously deemed

- ▶ **Summary** of this Article
- ▶ **dEbate:** Submit a response to this article
- ▶ Similar articles found in:
SCIENCE Online
ISI Web of Science
- ▶ Search Medline for articles by:
Shirts, M. || **Pande, V. S.**
- ▶ Search for citing articles in:
ISI Web of Science (2)
- ▶ Alert me when:
new articles cite this article
- ▶ Download to Citation Manager
- ▶ Collections under which this article appears:
Biochemistry
- ▶ **Computers/Mathematics**

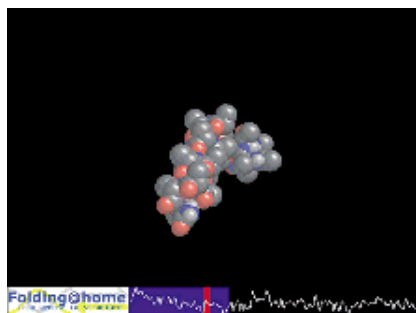
unparallelizable to be performed on hundreds or thousands of computers with very little communication between the processors.

Even if an algorithm can be parallelized, it may still be poorly suited for distributed computing. Consider, for example, simulations of the dynamics of biomolecules at the atomic level. Such simulations are traditionally limited to the nanosecond time scale. Duan and Kollman have demonstrated that traditional parallel molecular dynamics simulations can break the microsecond barrier (3), provided that one uses many tightly connected processors running on an expensive supercomputer for many months. This style of calculation requires, however, that the processors frequently communicate information and is thus poorly suited for worldwide distributed computing, where computer communication is thousands of times slower than the interprocessor communication in today's supercomputers.

Recently, an algorithm has been developed that helps address the problems of both parallelization and communication by allowing loosely connected multiple processors to be used for molecular dynamics (4, 5). The Folding@home project (5) has shown that this algorithm can reach orders of magnitude longer time scales than have previously been achieved when used for distributed atomistic biomolecular dynamics simulations. The design of similar algorithms for parallelization will likely play a major role in adapting other problems in computational biophysics (such as the design of more effective drugs) and other fields for distributed computing.

The ability to engage users to run the simulation software is central to the success of worldwide distributed computing. First, the user must have some interest in volunteering his or her computer. SETI@home and distributed.net have had great success in generating excitement about their projects. Biological and biomedical applications may have an even greater potential for generating public interest. Some commercial ventures even plan to expand this resource by paying users for their excess CPU time (6).

Second, distributed systems must not interfere with the user's personal use. This is most commonly (and perhaps most elegantly) done using screen savers (see the figure). The user downloads and installs the screen saver from the project's Web site. The vast majority of idle computer cycles will then be used for the project, without interfering with the user's work. To perform a calculation, the screen saver downloads some task from the project's server, performs the required calculation, returns the results to the server, and then repeats the cycle. To address networking and security issues, many projects use the same techniques as Web browsers and Web servers, because these methods of distributing data from client to server are well developed and secure. The project's server(s) must be carefully designed to handle the enormous number of clients in distributed computing.



Merging research and education. The Folding@home screen saver shows a graphical representation of the protein and its potential energy as it is folding, making the research more visually accessible to the public contributing to the project.

There are at least 300 million personal computers on the Internet (7). Up to 80 to 90% of their CPU power is wasted. If each distributed computing project involved 500,000 active users, as SETI@home currently claims, and half of all PCs now connected to the Internet participated, there would be sufficient capacity for 300 SETI-sized projects worldwide.

The world's supply of CPU time is very large, growing rapidly, and essentially untapped. Used to analyze the data generated by recent genomic and proteomic efforts or conduct other important calculations, distributed computing could raise biological and other scientific computation to fundamentally new predictive levels.

References and Notes

1. D. Butler, *Nature* **402**, C67 (1999); Netwatch, *Science* **289**, 503 (2000); Random Samples, *Science* **289**, 1135 (2000); B. Hayes, *Am. Sci.* **87**, 118 (1998) [American Scientist].
2. J. Kaiser, *Science* **282**, 839 (1998). See also <http://setiathome.ssl.berkeley.edu>
3. Y. Duan, P. A. Kollman, *Science* **282**, 740 (1998).
4. A. F. Voter, *Phys. Rev. B* **57**, 13985 (1998) [APS].
5. I. Baker *et al.*, in preparation. See also <http://foldingathome.stanford.edu>
6. D. Cohen, *New Sci.* **167** (no. 2247), 11 (2000). These ventures could potentially reduce availability of CPU time for noncommercial projects, although many are scientifically focused themselves and are operating in conjunction with academic researchers.
7. Number of computers on the internet estimated by using total PCs sold worldwide in the past 3 years (1998-2000), as published by IDC (www.idc.com).
8. The authors acknowledge S. Doniach for useful discussions and thank M. Levitt for a thorough review of our manuscript. Supported in part by the Fannie and John Hertz Foundation and the Stanford Graduate Fellowship program.

The authors are in the Department of Chemistry, Stanford University, Stanford, CA 94305-9450, USA. E-mail: pande@stanford.edu

- ▶ **Summary** of this Article
- ▶ **dEbate: Submit a response to this article**
- ▶ Similar articles found in:
SCIENCE Online
ISI Web of Science
- ▶ Search Medline for articles by:
Shirts, M. || **Pande**, V. S.
- ▶ Search for citing articles in:
ISI Web of Science (2)
- ▶ Alert me when:
new articles cite this article
- ▶ Download to Citation

Manager

- ▶ Collections under which this article appears:
Biochemistry
- ▶ **Computers/Mathematics**

Volume 290, Number 5498, Issue of 8 Dec 2000, pp. 1903-1904.
Copyright © 2000 by The American Association for the Advancement of Science.



▲ PAGE TOP