

Gaming Tech Tackles Alignment

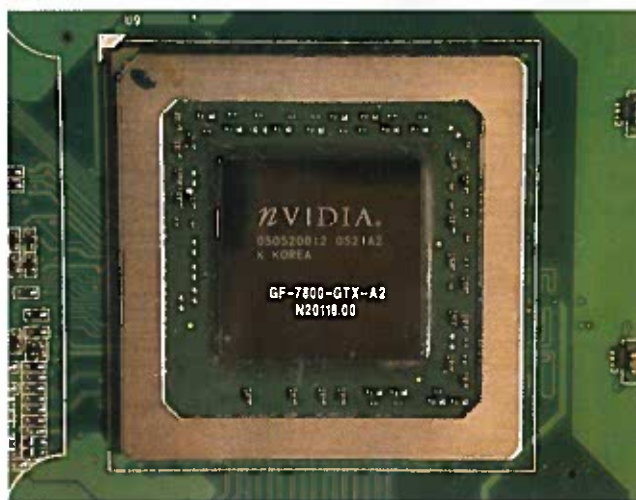
Tired of waiting for your Blast results? GPUs boost the algorithm's speed at a fraction of the cost of an FPGA — and you probably already have one.

By Jennifer Crebs

There's really nothing new about the problem: as sequence capacity grows, so does the need for high-performance computing tools to scan through rapidly growing databases. To answer this call, a few researchers are proposing a novel solution: use graphics hardware to facilitate better, faster alignments.

In recent years, computer scientists in a relatively new field known as GP-GPU programming, or general purpose computing on graphics processing units, have started putting GPUs to work on tasks outside of the domain for which they were originally designed. Only in the last year have preliminary results come out suggesting that the same graphics card built to satiate consumers' appetite for sophisticated video games may be put to good use in bioinformatics.

Wayne Huang, who works for both the DOE's Joint Genome Institute and Lawrence Livermore National Laboratory, presented initial results on his group's work to speed up sequence comparisons at Cold Spring Harbor's Genome Informatics conference late last year. Also last year, Bertil Schmidt, an assistant professor at Nanyang Technological University in Singapore, presented a paper on similar work at the IEEE Tencon Australia meeting. Both Huang and Schmidt have generated



striking results in accelerating sequence alignments by implementing a dynamic programming algorithm on graphics cards, thereby achieving speedups ranging from six- to 10-fold over standard CPUs.

Compute-intensive search heuristics based on the Smith-Waterman dynamic programming algorithm, such as Blast, are typically performed with run-of-the-mill CPUs. They can be significantly hurried up with hardware accelerators like the field-programmable gate array, but those aren't cheap — current FPGA systems can cost tens of thousands of dollars — and personalizing such setups requires hardware design skills to customize configurations to algorithms. This is “generally more complicated than writing new code for a programmable architecture” like that featured on graphics cards, Schmidt says.

Unlike FPGA-based accelerators, commodity graphics cards are available at low prices. The NVidia GeForce 6800 GT card that Huang put to use in initial tests runs for about \$400 and can achieve a six-fold speedup over CPUs, such as the 2.0 GHz AMD Opteron. For a few hundred dollars more, Schmidt says that one could accelerate algorithms even more using the NVidia 7800 GTX card. “The GeForce 7800 has a peak performance of 165 gigaflops, and the price is around \$600 to \$700. Compared with

the latest Pentium IV that has 15 gigaflops, the peak performance is 10 times higher,” says Schmidt.

However, the GPU implementation isn't as fast as FPGA-based systems. Huang's tests have shown that Time-Logic's DeCypher Engine G4 clocked in at speeds 15 to 20 times faster than the GPU version. Schmidt also allows that, while GPUs can achieve a 10-fold speedup over CPUs, FPGAs can run through the same query 100 times faster.

That said, high-end graphics cards are cheap and accessible. “Many labs already have the hardware needed to accelerate Smith-Waterman alignments,” says Huang, pointing out that clusters and workstations are usually outfitted with the necessary hardware. GPUs thus represent a “zero-cost solution” for many labs, according to Schmidt.

In order to use GPUs for sequence analysis, both Huang and Schmidt created approaches that reformulate Smith-Waterman to take advantage of the architecture used by graphics hardware. GPU-based computation requires a fixed order of processing stages, which is supported by multiple processors working in parallel. This graphics pipeline does not allow for the same flexibility in programming that you see in a CPU, Schmidt says, as code must be tailored to processors executing the same function concurrently.

Huang also cites the limitations inherent in GPU architecture as a challenge in programming graphics cards for sequence comparisons. But once achieved, GPU-based solutions are "relatively simple to implement," he says. Huang has a paper detailing how this works, and he expects that developers familiar with the OpenGL Shading Lan-

guage, the programming language used to prime the cards for sequence analysis, will be able to get started without much trouble. Schmidt adds that implementation requires "a relatively new graphics card, as the older ones do not support this kind of programmability."

But the programming capacity of a GPU could very well improve in future versions of the technology. Just as the FPGA industry has its roots in the mobile communications market, Schmidt notes, the latest generation of sophisticated graphics cards owes its existence to increasing entertainment and gaming demands. He expects the evolution of graphics hardware to continue at a rapid pace, thanks to the mounting demands of the video game market. "The gaming industry needs more sophisticated lighting and shading effects," Schmidt says, adding that features like enhanced programmability

have been increasing at a rate of up to two and a half times per year.

Huang's team is currently in the process of developing an API that can detect whether a supported GPU is available in a given compute system, which would "considerably accelerate the [alignment] process." If a supported graphics card is available, the API will accelerate Smith-Waterman automatically; if not, the alignment will be performed in the software that's requested. Another future goal: both Schmidt and Huang hope to make the implementation available as a grid resource.

The programming models proposed by both Schmidt and Huang are versatile, and are not exclusive to GPUs. Both researchers point out that the models may also be extended to other bioinformatics algorithms, and could be implemented on platforms such as a Cell processor in future applications. **GT**



ARDAS™

Track - Organize - Analyze

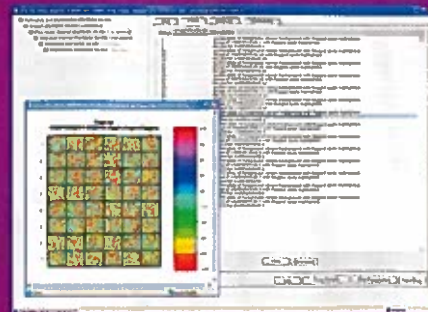
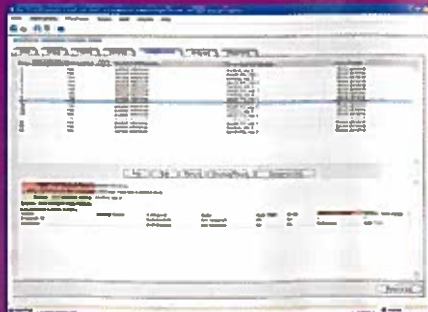
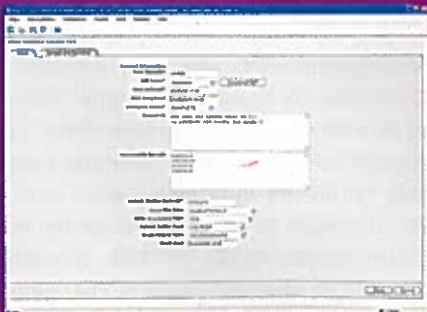
**2005
Quality Award
Winner**

The Array Repository and Data Analysis System provides a complete and fully integrated enterprise solution for **microarray** data acquisition, management, and analysis

LIMS

Warehouse

Analysis



Academic \$1900
Commercial \$2900



**For more information
visit: www.3rdmill.com/ardas
email: sales@3rdmill.com**