

COMPUTING A CURE



Lenovo partners with the Ontario Cancer Institute to help accelerate cancer research

Cancer. The very mention of the disease brings on feelings of dread. It's the leading cause of death worldwide. Yet, most cancers can be treated and survival rates have climbed for many forms of cancer thanks to extensive research and the development of more effective treatments. To improve our understanding of cancer and its treatment, scientists are leveraging new technologies to diagnose the disease in its earliest stages.

Lenovo is working with a group of researchers at the Ontario Cancer Institute (OCI) at Princess Margaret Hospital in Toronto, Canada that is studying proteins to improve our understanding of their function in cancer. Led by Dr. Igor Jurisica, this team is using Lenovo ThinkStations to analyze millions of images of proteins and compiling a database from which it hopes to be able to gain new insights into the mechanisms of cancer and other complex diseases that will lead to improved diagnostics and treatments.

lenovo FOR
THOSE
WHO DO.



Dr. Igor Jurisica leads a team of scientists at the Ontario Cancer Institute.

"We need to understand the structure of the proteins in order to understand the diseased and healthy states," said Jurisica. "And we also need the three-dimensional structure if we want to design new drugs to change the function of these proteins." That 3D structure is determined using protein crystallography, essentially a form of very high-resolution microscopy in which a beam of X-rays strikes a crystal and diffracts into many specific directions. A three-dimensional image can then be compiled from the angles and intensities of these diffracted beams.

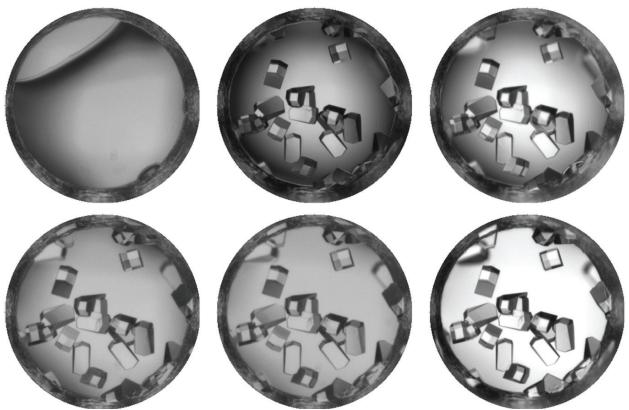
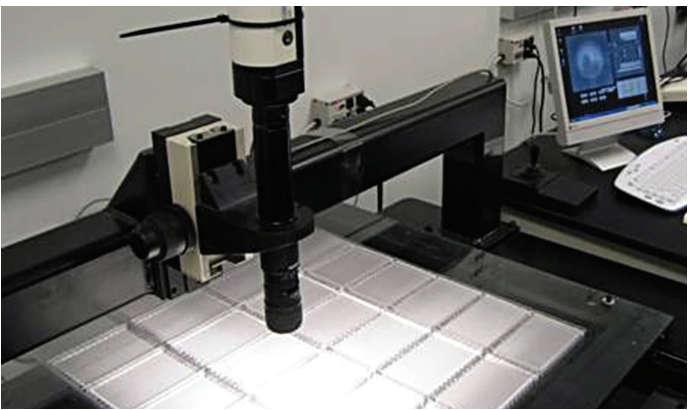
HOW TO ANALYZE MILLIONS OF IMAGES

In order to understand all of the variables, the researchers need to study a lot of images. Jurisica and his team work in conjunction with researchers at the Hauptman-Woodward Medical Research Institute (HWI) in Buffalo, New York. There, scientists use pipetting robots to grow the protein crystals for use in the study. According to Jurisica, the basic premise is quite simple, but his research quickly becomes complex due to the sheer numbers involved. "Every protein they run is tried in 1,536 different combinations of crystallization conditions at once," said Jurisica. But that's just the start. "We have been working with HWI for about 10 years. They've screened over 13,000 proteins and each protein is tested with 1,536 conditions." Then, the additional variable of time is introduced into the equation. Six images are made of each crystal as it grows over a period of one month. Jurisica did the math. "We have nearly 120 million images."

It's a brute force approach that is only possible through the use of highly

parallel computational analysis of the data. Since November 2007, Jurisica and his team have been running their project on the World Community Grid as part of a Help Conquer Cancer project. The World Community Grid's mission is "to create the world's largest public computing grid to tackle projects that benefit humanity." Similar to the SETI At Home project, which harnesses huge numbers of individual computers in a search for signs of extra-terrestrial intelligence, the World Community Grid enlists the aid of volunteers whose computers, when idle, download data, perform calculations, and return results back to the grid server. Jurisica's research is just one of nearly a dozen projects currently active on the grid. The World Community Grid runs on software developed at the University of California at Berkeley, with funding from the National Science Foundation. At present, the grid includes nearly 1.8 million individual computers worldwide and has compiled a total run time of more than 450,000 CPU years.

Since Jurisica's Help Conquer Cancer project began, more than 145 million



Protein crystals are grown in plates, each containing 1,536 individual wells. Each well in each plate is then photographed on a computer-controlled imaging station. An image is made of each successful protein result at six different time points, to record the changing crystallization conditions.

results have been returned. "We get about 54 CPU years a day worth of computing," Jurisica noted. He expects to complete the analysis of his current set of data by the end of this year. But that's just the start. Now, Jurisica and his team will use all of the data they've already collected to learn the principles that link the crystalline results to properties of the proteins and use that knowledge to improve the numbers of successfully crystallized proteins, and in turn to zero in on potential cures. "We can start additional analysis that was not possible before because such detailed results on 13,000 proteins didn't exist, and we did not have computing power available on the World Community Grid to comprehensively analyze these images."

One goal is to move the analysis from batch processing to real-time processing. But to do that, he can no longer rely solely on the grid of far flung computers. His colleagues in Buffalo continue to screen 200 to 300 proteins each month. "To continue to do this kind of analysis we need a platform where we can bring it into the lab," Jurisica continued, "and do it hopefully close to real-time."

ANALYSIS COMES TO THE WORKSTATION

To accomplish that real-time analysis, Jurisica has turned to Lenovo. Using Lenovo ThinkStation D20 workstations equipped with NVIDIA Quadro and Tesla GPU computing technologies, the OCI team has gained an enormous performance advantage by turning the original program into a GPU-enabled version. "The numbers that we are getting from this evaluation really start to be exciting because they are about 65 times faster than what we are getting on the grid per image!" noted Jurisica.

He explained that the analysis of a single image on a single CPU takes

about 45 minutes to return a result. But when the same analysis is run on the local Lenovo ThinkStation workstation with an NVIDIA Tesla C2050 GPU computing processor, the time shrinks to just one or two minutes. Running the calculations on the system's NVIDIA graphics processing unit (GPU) rather than on the traditional CPU has much to do with that improvement.

GPU computing uses the graphics processing unit in conjunction with the CPU. The sequential part of Jurisica's application runs on the CPU and the computationally-intensive part, analyzing the data contained in the protein crystallography images, is accelerated by the NVIDIA GPUs.

Jurisica initially chose Lenovo based on his past experience with IBM, having used IBM systems for various database and calculation functions and an IBM HS21 Linux cluster with 1,344 CPUs for some of the team's calculations. "But in terms of moving into more research-oriented and interactive computation, we needed workstations that could handle complex calculations and high-end graphics output," said Jurisica.

The team has also developed a close relationship with NVIDIA. Jurisica has experimented with several different NVIDIA graphics accelerators, including the Quadro FX 3800, Quadro 4000, Quadro 5000, Quadro 6000, and Tesla C2050. Again, he had some previous experience working with NVIDIA, having used the company's cards for other visualization projects, but "had never really pushed the GPU aspect," he says.

PIONEERS IN PARALLEL GPU COMPUTING

One large benefit of working with Lenovo and NVIDIA were the companies' combined depth of experience. NVIDIA was a pioneer in the development of GPU computing and Lenovo was

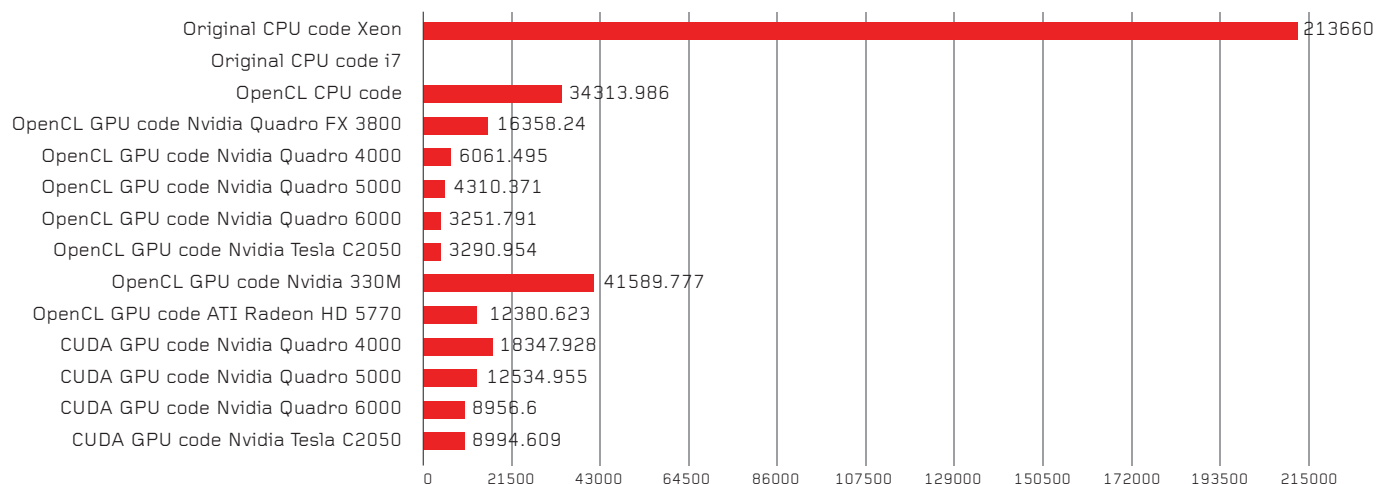


By running his analysis locally on Lenovo ThinkStation D20 workstations, Dr. Jurisica and his team are able to analyze images 65 times faster.

the first workstation vendor to offer configurations that included Tesla high-performance computing (HPC) solutions. NVIDIA's efforts evolved over the years with the company's introduction of its massively parallel CUDA, which consists of hundreds of processor cores inside the graphics card that work together to crunch through data. At the same time, NVIDIA developed the CUDA parallel programming model that helps facilitate the creation of programs designed to run on the GPU. While there are other evolving standards for GPU computing, such as OpenCL, Jurisica has found that NVIDIA and CUDA currently yield the best performance and the Lenovo ThinkStations provide a highly reliable platform for his work.

Jurisica also makes use of NVIDIA's Scalable Link Interface or SLI, which enables him to link multiple graphics cards together to further increase the number of parallel processing cores that can be harnessed for his calculations. The Lenovo ThinkStation D20 workstations are equipped with a pair of NVIDIA graphics boards.

Total Running Times for 50 Test Images (sec)



Jurisica and the team at the Ontario Cancer Institute achieved dramatic performance improvements by harnessing multiple GPUs in a single Lenovo ThinkStation workstation.

One board powers the display but both contribute to the calculations. A pair of NVIDIA Quadro 6000 or Tesla C2050 GPUs in the Lenovo workstation results in 896 CUDA cores all working simultaneously to analyze individual crystallization images.

Although past experience may have drawn Jurisica to Lenovo and NVIDIA, he quickly found justification for his decision. "There are not that many companies that are truly interested in working with researchers to push the envelope," said Jurisica. But he has found that willingness to be alive and well at both companies. "They are interested in hearing what problems we encounter," he noted, and then working with his team to solve those problems, even giving them access to new technologies that

have not yet been released to the market. "I appreciate this research aspect, where we can explore new technologies earlier," he said.

"If the grid would not be around, we would not be computing this comprehensive analysis on this large number of images, because we simply would not be able to afford to wait a couple hundred years for the results," Jurisica said. When it came time to move his research to local workstations, representatives at Lenovo and NVIDIA worked closely with Jurisica and his team to determine the best workstation and graphics cards to meet their unique needs. Since then, both Lenovo and NVIDIA have continued to be very responsive to the needs of the cancer research team. "When we had questions about some of the optimization issues

and really going deeply into the code optimization, NVIDIA and Lenovo experts were readily available and interested to work with us on this," Jurisica noted.

Like Jurisica's initial brute force approach to analyzing all those millions of images, he could have eventually determined the optimum configuration of workstation and graphics card. But thanks to the responsiveness of Lenovo and NVIDIA, such trial and error was unnecessary. "When you can ask an expert, then you have more brains and more experience helping solve the problem," said Jurisica. And when the problem is finding a cure for cancer, the researchers at the Ontario Cancer Institute really appreciate having Lenovo and NVIDIA as part of their team.