September 2011 Update

Milestones

As of October 1, 2011, 204,067,016 results were returned. Thank you all for your continued support and contribution – in average over 54 CPU/years a day.

We are finalizing a new proposal – Help Conquer Cancer II – focusing on systematically identifying molecular markers that would improve early diagnosis and prognosis prediction for multiple cancers.

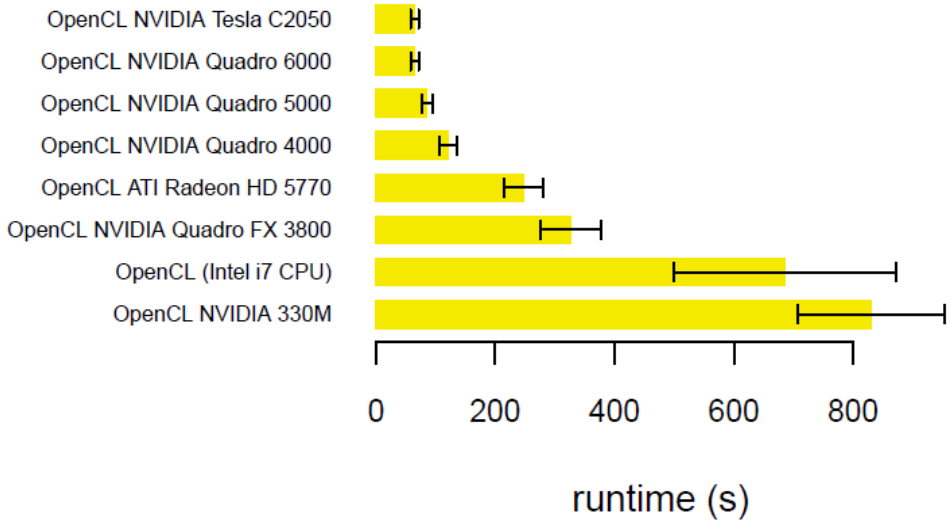
Beta version of GPU-accelerated HCC code

We have developed a GPU version of an automated image analysis and classification system for protein crystallization trial images from the Hauptman Woodward Institute's High-Throughput Screening lab. The analysis step computes 12,375 numerical features per image. Using these features, we have trained a classifier that distinguishes 11 different crystallization outcomes, recognizing 80% of all crystals, 94% of clear drops, 94% of precipitates.

We focused our efforts on optimizing the GLCM features, which consume 98% of the run-time on the CPU version. We chose the OpenCL 1.0 language due to its support for multiple vendors and device types.

To prepare for a broader WCG run, we have considered NVIDIA and ATI implementations for Linux and Mac OS. Theoretically, a program written in accordance to the OpenCL API specifications must run correctly on all supporting devices. Obviously this does not guarantee optimal performance due to the differences in hardware architectures. For those reasons we ended up writing separate versions of kernel code for CPU, NVIDIA and ATI cards.

The results are exciting. While CPU runtime averages 4092 seconds (single threaded) on an Intel Xeon, but only 65 seconds on an NVIDIA Tesla C2050.



We were able to achieve a significant performance improvement for this problem. Our final version of OpenCL code runs 6 to 60 times faster compared to the original optimized C++ code. It took over 3 days of CPU time to process our small 50 image test set, and under 2 hours on the NVIDIA Tesla C2050 GPU.

There is always a tradeoff between precision and performance on GPU. We first computed relative error of the OpenCL results vs. CPU version**.** However, improving performance alone is not sufficient – we need high accuracy of the results. To evaluate our OpenCL implementation we used the existing optimized C++ code that also runs on the World Community Grid. Almost 97% of values are correct up

to four significant digits. But ultimately, the most important is whether the classifier accuracy is negatively affected by this loss of precision. To verify that any shortcuts we took in the OpenCL version would not affect our classification, we tested final code on a set of 4,500-image subset of our image-classifier's training/testing set. We then applied our classifier to both GPU- and CPU-computed image features and found only eight images that were assigned different classes. All eight of these images belonged to an ambiguous category (human experts disagreed on the image's true class), suggesting that discrepancies between GPU and CPU features do not pose any problems.

The GPU version therefore runs quickly enough to allow same-day analysis of crystallization-trial images, and correctly enough to run alongside the CPU version on the World Community Grid. We expect GPU-enabled Help Conquer Cancer to enter beta testing in the near future.

These results were recently published in:

> Kotseruba, Y., Cumbaa, C. A., Jurisica, I**.** High-throughput protein crystallization on the World Community Grid and the GPU. *J Physics: Conference Series*, High Performance Computing Symposium (HPCS 2011: HPC in Medical Science), 2011. In press

Other progress

Other progress in our research is highlighted in a newly launched web site for the Cancer Gene Encyclopedia project: http://ophid.utoronto.ca/cgep. We have released version 1 of our portal for microRNA:gene target prediction – mirDIP – http://ophid.utoronto.ca/mirDIP.

We have recently teamed up with Dr. J. Dick and through this research identified stem cell signature with prognostic value for leukemia patients:

> Eppert, K., Takenaka, K., Lechman, E.R., Waldron, L., Nilsson, B., van Galen, P., Metzeler, K., Poeppl, A., Ling, V., Beyene, J., Canty, A.J., Danska, J.S., Bohlander, S.K., Buske, C., Minden, M.D., Golub, T.R., Jurisica, I., Ebert, B.L., Dick, J.E. Stem cell gene expression programs influence clinical outcome in human leukemia, *Nat Medicine*, **14**(8): 822–827, 2011.

Another exciting new result in head & neck cancer has a potential to significantly modify how individual patients are treated. The study was also highlighted in the BMC press release:

> A gene signature in histologically normal surgical margins is predictive of oral carcinoma recurrence. Reis, P.P., Waldron, L., Perez-Ordonez, B., Pintilie, M., Natalie Galloni, N., Xuan, Y., Cervigne, N.K., Warner, G.C., Makitie, A.A., Simpson, C., Goldstein, D., Brown, D., Gilbert, R., Gullane, P., Irish, J., Jurisica, I., and Kamel-Reid, S., *BMC Cancer*, 2011. In press.

Summary of other publications can be found at http://www.cs.utoronto.ca/~juris/publications.htm.

And for those in love of cycling – we have completed our 4th Ride to Conquer Cancer in Ontario and the first one in Quebec. Please, consider joining us or supporting us. You can find more information on http://www.cs.toronto.edu/~juris/gallery.htm.

Thank you,

C. A. Cumbaa, Yulia Kotseruba and I. Jurisica