

Vorwort zum Workshop “Massively Parallel Computational Biology on GPUs”

Prof. Dr. rer. nat. Kay Hamacher (Bioinformatics & Theo. Biology, TU Darmstadt)
Prof. Dr.-Ing. Michael Goesele (GRIS, TU Darmstadt)

Die Lebenswissenschaften befinden sich derzeit in einem historischen Umbruch. Nach der Sequenzierung von mehr als 700 Genomen, einer exponentiell steigenden Anzahl molekularer Strukturen und unzähligen Studien (zum Beispiel zu biochemischen Details), wird der Bedarf für eine integrative, theoretische Behandlung dieser Experimentaldaten immer offensichtlicher:

“The history of science is emphatic: a discipline usually matures [...] after it has developed mathematical theory”

P. Turchin, Professor for Ecology and Evolutionary
Biology, University of Connecticut

Da die Biologie und verwandte Disziplinen sich dadurch auszeichnen, dass einerseits Prozesse und Phänomene auf verschiedenen Längen- und Zeitskalen stattfinden und andererseits große Datenmenge herangezogen werden können und müssen, ist der Einsatz von Rechnern imperativ. Allerdings steigen auch die Anforderungen an Berechnungen und Simulationen massiv, wenn diese unterschiedlichen Skalen integriert werden sollen. In vielen Fällen sind daher nur noch Parallelrechner in der Lage, für das Experiment sinnvolle Antworten zu geben.

Um die dadurch explodierenden Kosten zu verringern, bietet sich seit einiger Zeit die Nutzung von Graphikkarten (Graphical Processing Units, GPUs) zur Berechnung an. Diese GPUs zeichnen sich durch eine hohe Leistungsdichte (zur Zeit etwa 1 Teraflops pro GPU), eine schnelle Steigerung der Leistung verglichen mit herkömmlichen CPUs, und ein herausragendes Preis-/Leistungsverhältnis aus. Dies ermöglicht ganz neue Anwendungen beziehungsweise die Betrachtung komplexer Systeme, die bisher aufgrund des Rechenaufwands unerreichbar waren. Aktuelle Implementierungen beschäftigen sich auch in der Tat mit einer Vielzahl von Fragestellungen im Bereich der Computational Biology: So sind erste Studien publiziert zu verbesserten Multiple-Sequence-Alignments auf GPUs, der Vorhersage von RNA-Sekundärstrukturen mittels massiv-paralleler Suche, oder auch die Übertragung von Molekulardynamik-Simulationen auf GPUs (namentlich in Form von NAMD und HOOMD).

Von Seiten der reinen Informatik bieten die genannten Problemstellungen ein vielversprechendes Szenario zur Entwicklung neuer massiv-paralleler Programmieransätze. Dies schließt sowohl die Programmierung aktueller Graphikkarten (bekannt unter dem Namen General Purpose Computation on GPUs, GPGPU) als auch zukünftiger Many-Core-Architekturen mit ein.

Ziele und Programm

All dies zeigt die derzeit vorherrschende Dynamik auf diesem Feld, dem von verschiedenen Seiten ein hohes Potential beigemessen wird. Wir sind davon überzeugt, dass die Verfügbarkeit von entsprechender Software das Feld der Computational Biology aber auch der restlichen Computational Sciences in den nächsten Jahren grundlegend verändern wird. Daher war es uns wichtig, dass dieses Feld auch im Rahmen der diesjährigen Jahrestagung der Gesellschaft für Informatik unter dem Motto “Im Focus das Leben” repräsentiert ist. Ziel dieses Workshops war es daher einerseits, Forscher, die sich bereits mit massiv-parallelen Berechnungen auf GPUs beschäftigen, zusammen zu bringen und einen Austausch zu ermöglichen. Andererseits sollten auch Gruppen, die bisher “GPU-fern” sind, an diese neuen Ansätze herangeführt werden.

Das Workshop-Programm gliedert sich in folgende Teile:

- Ein eingeladenes Tutorial zum Thema massiv-parallele Berechnungen auf aktuellen GPUs. Das Tutorial wird von Gernot Ziegler (NVIDIA) gehalten und gibt eine generelle Einführung in diesen Themenbereich. Es dürfte somit auch für Zuhörer aus Bereichen jenseits der Computational Biology sehr interessant sein.
- Ein eingeladener Vortrag mit dem Titel “Scientific Computing on Multi-GPU Systems.” Dieser Vortrag von Robert Strzodka (MPI Informatik) beschäftigt sich insbesondere mit den Herausforderungen heterogener Systeme, die parallele Rechner oder Cluster mit massiv-parallelen GPUs kombinieren.
- Durch Review ausgewählte Beiträge zu massiv-parallelen Verfahren in der Computational Biology.

Besonders danken möchten wir an dieser Stelle den Vortragenden der eingeladenen Beiträge und den Mitgliedern des Programmkomitees, ohne deren Unterstützung wir den Workshop in dieser Form nicht hätten gestalten können:

- Justin Gullingsrud, D. E. Shaw
- Volkhard Helms, Universität des Saarlandes
- Mireille Krier, Merck KGaA
- David Luebke, NVIDIA
- Wolfgang Müller-Wittig, CamTech, Singapur
- Robert Strzodka, MPI Informatik
- Wolfgang Wenzel, FZ Karlsruhe/KIT

*Michael Goesele und Kay Hamacher
Darmstadt, im Juli 2009*