Simulationen auf Grafikkarten: vom Videospiel zur Materialforschung

Von Martin Oettel und Peter Virnau

Nachwuchswissenschaftler des Instituts für Physik erschließen die Rechenleistung von Grafikkarten, wie sie bei Computerspielen zum Einsatz kommen, für materialwissenschaftliche Forschungen. Aufwendige numerische Rechnungen und Simulationen lassen sich auf diese Weise bis zu hundertfach beschleunigen.

Die Erforschung von Materialeigenschaften ist zunehmend von einem massiven Computereinsatz geprägt. Mittlerweile werden immer leistungsfähigere Rechner eingesetzt, um das Verständnis makroskopischer Materialeigenschaften aus dem Zusammenspiel der mikroskopischen Bausteine, der Moleküle, zu erhöhen. Das detaillierte Studium der Eigenschaften und Wechselwirkungen von Molekülen kann zwar prinzipiell auch mittels quantenchemischer Methoden präzise durchgeführt werden (vgl. den Beitrag von Gregor Diezemann und Kollegen). Die Beschreibung eines Ensembles von Molekülen mit diesen Methoden und heute zugänglichen Computern bleibt jedoch in der Regel auf Phänomene beschränkt, die auf sehr kleinen Zeit- und Längenskalen ablaufen. Skalen, die die Wechselwirkungen zwischen den Molekülen in einem Material bestimmen, sind hingegen fast immer deutlich von den entsprechenden quantenmechanischen Skalen getrennt. Hieraus entsteht ein „vergröbertes" Bild des Materials, welches Atome oder gar Moleküle als klassische Newtonsche Teilchen beschreibt, die durch Kraftfelder miteinander in Wechselwirkung treten. Die Idee der Vergröberung kann aber auch zu noch langsameren Zeitskalen und größeren Längenskalen fortgesetzt werden, wenn die eigentlich wichtigen mikroskopischen Bausteine eines Materials bereits aus vielen Molekülen zusammengesetzt sind, wie es zum Beispiel bei Polymeren, Membranen oder kolloidalen Objekten der Fall ist. Die Verknüpfung dieser verschiedenen Ebenen führt zu Multiskalen-Modellen, wie sie in der Theoriegruppe des Max-Planck-Instituts für Polymerforschung zum Einsatz kommen (vgl. den Beitrag von Christine Peter und Kurt Kremer).

Traditionell werden umfangreiche Molekulardynamik-Simulationen, die auf den Newtonschen Bewegungsgleichungen basieren, auf Parallelrechnern mit Hunderten von Prozessor-Kernen durchgeführt. Die Notwendigkeit zur Parallelisierung der Simulationsmethoden ist auch dadurch bedingt, dass die Erhöhung der Rechenleistung der Computer im letzten Jahrzehnt eher durch die Entwicklung paralleler Architekturen auf den Prozessoren (CPUs) erreicht wurde als durch die Erhöhung der Leistungsfähigkeit eines einzelnen CPU-Kernes. Darüber hinaus ist die Erhöhung der Rechenleistung über massive Parallelisierung im letzten Jahrzehnt auch an einem anderen Baustein moderner Computer konsequent vorangetrieben worden: an den Grafikkarten (GPUs – graphical processing units), wie sie etwa zur Beschleunigung von Videospielen zum Einsatz kommen. Der expandierende Markt für Computerspiele führte dabei zur Entwicklung immer leistungsfähigerer Grafikkarten, die eine immer realistischere Darstellung in hohen Auflösungen gewährleisten. Grafikkarten müssen dabei immer wiederkehrende, relativ einfache Fließkommarechnungen (gemessen in Flop/s – Rechenoperationen pro Sekunde) in hoher Geschwindigkeit durchführen. Dafür wurde eine parallele Architektur entwickelt, die – zum Beispiel bei einer modernen Grafikkarte der Firma NVIDIA – aus 30 Multiprozessoren mit jeweils acht einfachen Rechenkernen, in der Summe also aus 240 Kernen, besteht (im Vergleich zu vier Kernen auf modernen QuadCore-Prozessoren). Abbildung 1 zeigt eine solche Karte, die zwei der oben beschriebenen GPU-Einheiten umfasst. Werden diese GPU-Kerne gleichmäßig ausgelastet, so sind erheblich mehr Rechenoperationen möglich als auf herkömmlichen CPUs. Verdeutlicht man sich diese Leistungsunterschiede von GPU und CPU, so scheint es naheliegend, sich diese Kapazitäten auch im wissenschaftlichen Rechnen zunutze zu machen. Dabei ist zu berücksichtigen, dass sich nicht jedes Problem auf dieser „Architektur" effizient darstellen lässt. Lediglich Fragestellungen, bei denen immer gleiche (voneinander unabhängige) Rechenoperationen auf unterschiedliche Daten angewendet werden, eignen sich für eine solche Plattform. Bis vor wenigen Jah-

Abb. 1: Grafikkarte der aktuellen Generation (NVIDIA GTX 295).

© Cornelia Kirch / Institut für Physik / JGU Mainz

ren war es auch nicht möglich, Grafikkarten in einer höheren Programmiersprache anzusprechen, was aus Gründen der Portabilität und Programmierfreundlichkeit natürlich wünschenswert ist. Seit 2003 gibt es nun Bestrebungen, Programmierumgebungen für Grafikkarten zu entwickeln, die der weitverbreiteten Programmiersprache C angepasst sind. Hervorzuheben ist hier CUDA (Compute Unified Device Architecture) von NVIDIA, welches 2006 eingeführt wurde und sich seitdem steigender Beliebtheit auch für wissenschaftliche Anwendungszwecke erfreut.

Im Rahmen des Schwerpunkts für „Rechnergestützte Forschungsmethoden in den Naturwissenschaften" (SRFN) werden zurzeit in mehreren Mainzer Arbeitsgruppen fächerübergreifend Anwendungsmöglichkeiten von Grafikkarten für wissenschaftliche Fragestellungen getestet, beispielsweise bei Packproblemen in der Informatik (vgl. den Beitrag von Johannes J. Schneider und Elmar Schömer) oder für Berechnungen in der Hochenergiephysik. Unsere Arbeitsgruppen beschäftigen sich hingegen schwerpunktmäßig mit materialwissenschaftlichen Fragestellungen. Erste Erfahrungen mit dieser neuen Plattform wurden jedoch auch auf anderen Gebieten gesammelt. So wurden von Tobias Preis, einem Junior-Mitglied der Gutenberg-Akademie und Doktoranden unseres Institutes, Zeitreihenanalysen von Börsendaten auf der Grafikkarte implementiert.[1] Die enorme Rechenleistung ermöglicht hierbei, eingehende Informationen auf Millisekundenbasis ohne Einsatz eines Supercomputers in Echtzeit zu analysieren. Einer späteren Integration in ein Real-Time-Risikomanagementsystem ist ebenfalls möglich. Darüber hinaus wurde die Berechnungszeit für ein klassisches Modell des Magnetismus, das sogenannte Ising-Modell, im Vergleich zu einem einzelnen CPU-Kern um bis auf das Sechzigfache herabgesetzt.[2] Hierbei kam der sogenannte Schachbrett-Algorithmus zum Einsatz, der bereits seit Jahrzehnten bei der Parallelisierung auf gewöhnlichen Supercomputern angewendet wird. Man darf jedoch nicht vergessen, dass die Übertragung bekannter und bewährter Algorithmen auf die Grafikkarte nicht frei von Problemen ist. Denn es ist dafür zu sorgen, dass alle GPU-Kerne möglichst gleichmäßig „beschäftigt" werden, was gelegentlich sehr spezifische Detailarbeit an einem Programmcode erfordert.

Seit Ende 2008 wird auch vom SRFN ein Dissertationsprojekt mit einer Anschubfinanzierung gefördert, in dem ein Beitrag zu aktuellen Fragen der Materialforschung geleistet werden soll.

Für weiche oder biologische Materialien ist eine präzise Behandlung mit Multiskalen-Modellen in der Simulation, wie sie anfangs skizziert wurden, nicht unbedingt einfach. Zudem ist ein generisches Verständnis dieser Materialien oft noch sehr rudimentär. Das heißt, es ist letztlich unklar, welche Art von mikroskopischen Korrelationen zu den oft ungewöhnlichen makroskopischen Eigenschaften solcher Materialien führt. Beispiele für solche ungewöhnlichen Eigenschaften sind Phasentrennungsphänomene, nicht-monotones Verhalten unter Scherung sowie Übergänge zu gel- und glasartigen Zuständen mit extrem verlangsamter Dynamik, deren charakteristische Zeitskala um viele Größenordnungen über der Zeitskala der molekularen Dynamik liegt. Kolloidale Flüssigkeiten, die aus mikroskopischen Bausteinen verschiedener Größe bestehen, sind typische weiche Materialien, die ein derart komplexes Verhalten zeigen. Die Modellierung solcher Systeme mit einfachen, „generischen" Wechselwirkungspotentialen erlaubt sowohl eine Behandlung mit den üblichen molekulardynamischen Simulationsmethoden als auch mit Methoden der klassischen Dichtefunktionaltheorie. Konkret geht es in einem ersten Schritt darum, an einem einfachen Modell für eine kolloidale Flüssigkeit (Kugeln in einer Suspension) die mikroskopische Dynamik eines Testteilchens zu studieren. Im Nichtgleichgewicht enthalten diese dynamischen Korrelationen – als Funktion einer äußeren treibenden Kraft und der Zeit – Informationen über die Fließeigenschaften der Flüssigkeit und können auch das Auftreten von Übergängen zu glas- oder gelartigen Zuständen signalisieren. Dies ist ein Beispiel für den bereits erwähnten Zusammenhang zwischen mikroskopischen Korrelationen und makroskopischen Eigenschaften des Materials.

Quelle: 3

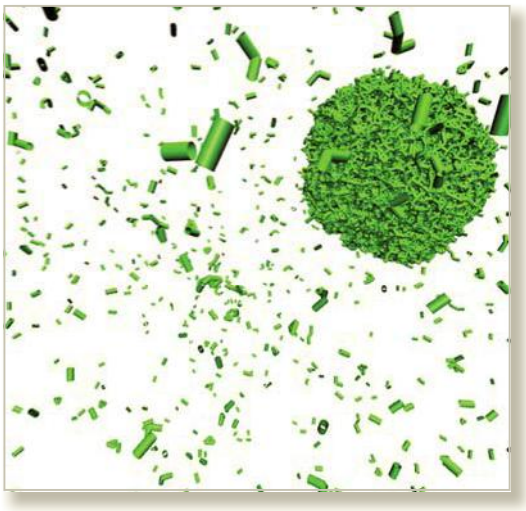

Abb. 2: Keimbildungssimulation mit HOOMD (Highly Optimized Object-oriented Molecular Dynamics); Endkonfiguration eines nukleierten Tröpfchens. In dieser Darstellung sind nahe beieinanderliegende Teilchen durch zylinderförmige Brücken verbunden.

Quelle: www.nvidia.com

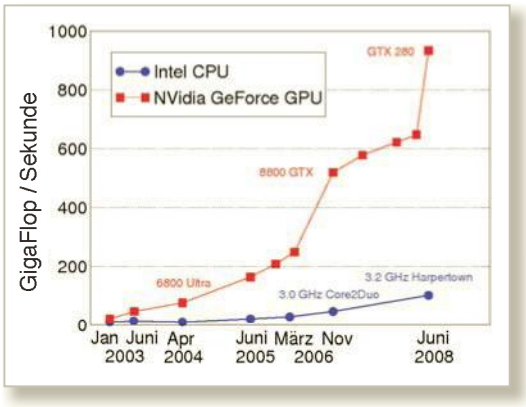

Abb. 3: Vergleich der theoretisch erreichbaren Rechenleistung von GPU und CPU in GFlop/s (engl.: Giga-floating-point-operations per second = Milliarden Rechenoperationen pro Sekunde).

Erste Testrechnungen mittels klassischer Dichtefunktionaltheorie auf Grafikkarten signalisieren auch hier Geschwindigkeitszuwächse, wie sie schon in den vorhergehenden Studien erzielt worden sind. Darüber hinaus haben wir auch erste Testsimulationen zur Phasenseparationskinetik und Keimbildung (Abb. 2) mit bereits bestehenden Molekulardynamikpaketen auf Grafikkarten unternommen.[3]

Die hier beschriebenen Ansätze stellen lediglich einen ersten Schritt zu einer weiterführenden Nutzung dieser neuen Technologie im wissenschaftlichen Rechnen dar. Das enorme Potential der Grafikkarten lässt sich nicht zuletzt an der Entwicklung der Rechenleistung im Vergleich zu herkömmlichen CPUs in den letzten Jahren veranschaulichen (Abb. 3). Diese Entwicklung eröffnet für die Zukunft die Perspektive, neuartige, an große Systeme und lange Simulationszeiten geknüpfte Fragestellungen anzugehen.

Summary

Young scientists from the department of physics utilize the computing power of modern graphic cards, which are primarily used to accelerate video games, for research in Materials Science. Acceleration factors of up to 100 are achieved for complex numerical calculations and simulations. In the near future a new class of problems can be tackled which require large systems and long simulation times.

Literatur

1. Preis T, Virnau P, Paul W, Schneider JJ. Accelerated fluctuation analysis by graphic cards and complex pattern formation in econophysics. Eingereicht 2009.

2. Preis T, Virnau P, Paul W, Schneider JJ. GPU accelerated Monte Carlo simulations of the 2D and 3D Ising models. J Comp Phys 2009; 228: 4468.

3. Anderson JA, Lorenz CD, Travesset A. General purpose molecular dynamics simulations fully implemented on graphics processing units. J Comp Phys 2008; 227: 5342.

Cornelia Kirch / Institut für Physik / JGU Mainz

Dr. Martin Oettel

Martin Oettel, Jahrgang 1973, studierte Physik an der TU Clausthal, der State University of New York in Stony Brook und der Eberhard-Karls-Universität Tübingen. Promoviert hat er im Jahr 2000 in Tübingen mit einem Thema aus der Elementarteilchentheorie. Nach einem durch die Alexander von Humboldt-Stiftung unterstützten Postdoc-Aufenthalt in Adelaide (Australien) wechselte er 2002 an das Max-Planck-Institut für Metallforschung in Stuttgart. Seither ist sein Forschungsschwerpunkt in der Theorie der Flüssigkeiten und kolloidalen Systeme angesiedelt. Seit 2006 leitet er eine unabhängige Nachwuchsgruppe in Mainz innerhalb des Sonderforschungsbereiches TR6 („Kolloide in externen Feldern"), die sich schwerpunktmäßig mit Ordnungsphänomenen von Kolloiden an Grenzflächen beschäftigt.

Cornelia Kirch / Institut für Physik / JGU Mainz

Dr. Peter Virnau

Peter Virnau, Jahrgang 1976, studierte Physik an der Ruprecht-Karls-Universität Heidelberg und an der Louisiana State University in Baton Rouge (USA). Nach seinem Master-Abschluss promovierte er in der Arbeitsgruppe von Prof. Kurt Binder an der Johannes Gutenberg-Universität Mainz über Monte-Carlo-Simulationen von Polymeren. Nach einem dreijährigen Forschungsaufenthalt am Massachusetts Institute of Technology in Cambridge (USA) kehrte er zur Habilitation nach Mainz zurück. Neben den oben erwähnten Grafikkarten gehören auch Monte-Carlo-Simulationen von Kolloiden und Polymeren sowie Keimbildung und topologische Eigenschaften von Proteinen zu seinen Forschungsschwerpunkten. Für seine biophysikalischen Arbeiten wurde ihm 2007 der Kalkhof-Rose-Gedächtnispreis der Mainzer Akademie der Wissenschaften und der Literatur verliehen.

Kontakt

Dr. Martin Oettel
Johannes Gutenberg-Universität Mainz
Institut für Physik
D-55099 Mainz
Tel. +49 (0) 6131-39 23 645
Fax +49 (0) 6131-39 25 441
Email: oettelm@uni-mainz.de
www.cond-mat.physik.uni-mainz.de/~oettel