

# Multicore- und GPGPU-Architekturen

Manuel Danisch

04. November 2010

## **Zusammenfassung**

Ziel dieser Arbeit ist es, einen Überblick über aktuell verfügbare CPU- und GPGPU-Architekturen zu geben. Es soll außerdem anhand bald erscheinender Designs und anderer Faktoren eine generelle Tendenz der Weiterentwicklung gezeigt und eine Langzeitprognose erstellt werden.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>3</b>
1.1	Inhalt . . . . .	3
1.2	Begriffe . . . . .	3
1.3	In-order / Out-of-order execution . . . . .	4
1.4	Gegenüberstellung Multiprozessor / Multicore / Multithread . . . . .	5
1.5	Homogene / Heterogene Multicores . . . . .	6
1.6	Technologien zur Verbindung der Kerne . . . . .	6
<b>2</b>	<b>CPUs</b>	<b>8</b>
2.1	IBM PowerXCell 8i . . . . .	8
2.2	Intel . . . . .	9
2.3	AMD . . . . .	10
<b>3</b>	<b>GPGPUs</b>	<b>11</b>
3.1	Nvidia . . . . .	11
3.2	AMD . . . . .	12
3.3	Intel . . . . .	12
<b>4</b>	<b>Ausblick auf zukünftige Entwicklung</b>	<b>13</b>
4.1	APUs . . . . .	13
4.1.1	Intel SandyBridge . . . . .	13
4.1.2	AMD . . . . .	14
4.2	Langzeitprognose . . . . .	14

# Kapitel 1

## Einleitung

### 1.1 Inhalt

In dieser Arbeit werden aktuelle und kommende CPU- und GPGPU-Architekturen vorgestellt. Die Grundlage zum Verständnis der späteren Kapitel wird in diesem Kapitel gelegt, in dem zentrale Begriffe und Technologien eingeführt werden. Danach folgen die zwei Hauptteile mit auf dem Markt befindlichen Multicore-CPU und GPGPU-Prozessoren. Die Arbeit endet mit der Vorstellung der nächsten Chipgeneration(en) und mit einer Prognose für die weitere Entwicklung der Architekturen in der Zukunft.

### 1.2 Begriffe

Als erstes ist die Frage zu klären, welche Art von Parallelität eigentlich betrachtet wird. Dazu werden einige Begriffe aus der Flynn'schen Klassifikation definiert:

**Single-Instruction Single-Data** Eine Instruktion auf ein Datum, die klassische Arbeitsweise des von-Neumann-Rechners

**Single-Instruction Multiple-Data** Eine Instruktion auf mehrere Daten, z.B. in den SSE Rechenwerken einer CPU

**Single-Program Multiple-Data** Ein Programm auf verschiedene Daten, wie SIMD ohne Barrier-Synchronisation nach jedem Befehl

**Multiple-Instruction Multiple-Data** Mehrere Instruktionen auf mehrere Daten, z.B. verschiedene unabhängige Threads auf verschiedene Daten

MIMD ist dabei relativ uninteressant, da zwar verschiedene Programme parallel auf den verfügbaren Kernen laufen können, die Programme aber nicht mit der Zahl der Kerne skalieren. Das eigentliche Ziel, das man gerne erreichen möchte, ist SPMD, da dabei ein Programm ohne gegenseitige Synchronisation parallel auf allen verfügbaren Kernen ausgeführt wird und auf verschiedene Datenobjekte parallel zugreifen kann. Im Idealfall führt SPMD dadurch zu einer sehr effizienten Ausnutzung aller vorhandenen Kerne. Es ist noch anzumerken, dass oft nicht zwischen SIMD und SPMD unterschieden wird. Meistens ist bei Hardware nur von SIMD die Rede (SSE bedeutet z.B. Streaming SIMD Extensions).

Ein Feld, in dem SPMD oft gut genutzt werden kann, ist die Bearbeitung von Grafikdaten auf einer GPU, z.B. der Erstellung von 3D-Grafiken für Computerspiele oder sonstiger Grafikmanipulationen. Da GPUs daher von Anfang an bzgl. des Hardwaredesigns darauf optimiert wurden, sind heute bei Problemen, die sich mit SPMD gut lösen lassen, GPUs den CPUs in der Rechengeschwindigkeit oft weit überlegen.

### 1.3 In-order / Out-of-order execution

CPUs benutzen schon seit längerer Zeit Pipelines, um die Geschwindigkeit zu erhöhen. Dies bedeutet, dass z.B. pro Takt bei einer vierstufigen Pipeline die vier Phasen

- Befehl holen
- Befehl decodieren
- Befehl ausführen
- Ergebnis zurückschreiben

parallel ausgeführt werden. Während also ein Befehl decodiert wird, wird schon der nächste geladen. Nun kann es zu Konflikten kommen (Hazards), so dass die Verarbeitung des gerade geholten Befehls verzögert werden muss (Stalling), da der Befehl z.B. ein Ergebnis eines früheren benötigt, welches noch nicht vorliegt. Dieses Stalling verursacht Lücken in der Pipeline. Dadurch werden die Funktionseinheiten nicht optimal ausgelastet und der Durchsatz sinkt.[21]

Es kommt nun auf die CPU an, ob sie die Pipeline in einem solchen Fall trotzdem auslasten kann.

**In-order execution** Die CPU behält die Befehlsreihenfolge bei

**Out-of-order execution** Die CPU kann die Befehlsreihenfolge ändern

Wenn eine CPU nur In-order beherrscht, dann kann sie die Lücken in der Pipeline nicht mit anderen Befehlen füllen. Eine out-of-order CPU hingegen kann z.B. spätere Befehle, die keine neuen Hazards verursachen, nach vorne ziehen und statt dem verzögerten Befehl in die Pipeline schieben. Der Nachteil bei out-of-order ist, dass die Umsortier-Logik mehr Transistoren benötigt und damit sowohl die Komplexität als auch die Wärmeabgabe des Chips erhöht.[19]

## 1.4 Gegenüberstellung Multiprozessor / Multicore / Multithread

Für den Programmierer ist es bei den heute immer noch weit verbreiteten hardwarenahen Sprachen durchaus von Bedeutung, wo genau ein bestimmter Prozess ausgeführt wird. Im Großen und Ganzen läuft ein Prozess:

- auf verschiedenen Prozessoren,
- auf demselben Prozessor auf verschiedenen Kernen,
- auf demselben Kern mit mehreren Hardware-Threads

Die ersten beiden Punkte sind relativ selbsterklärend. Es ist nur zu beachten, dass verschiedene Prozessoren meistens jeweils einen eigenen Speicher ansprechen, während verschiedene Kerne auf demselben Prozessor alle auf einen gemeinsamen Speicher Zugriff haben.

Eine Sonderstellung nehmen dabei die Hardware-Threads ein. Diese Technik wurde im Massenmarkt von Intel beim Pentium4 unter dem Namen Hyperthreading eingeführt. Dabei meldet die CPU dem Betriebssystem pro physischem Kern einen weiteren virtuellen Kern, wodurch die CPU mit insgesamt doppelter Kernzahl erscheint. Dies wurde eingeführt, um die Auslastung der Rechenwerke zu erhöhen, ohne dass man eine höhere Verlustleistung oder größere Fertigungsumstellungen in Kauf nehmen musste. Das große 'Problem' des Pentium4 war nämlich seine sehr lange Pipeline (Intel hatte die Absicht, sehr hohe Taktraten zu erreichen), in der sich trotz des out-of-order Kern-Designs große Lücken befanden. Indem man einen zweiten Thread pro Kern zuließ, gab es mehrere Befehle, mit denen sich die Pipeline füllen ließ, ohne weitere Hazards zu verursachen. Dementsprechend gering war jedoch auch die Beschleunigung, die man mit Hyperthreading erreicht hat.[25]

## 1.5 Homogene / Heterogene Multicores

Ein weiterer wichtiger Punkt ist der grundlegende Aufbau von Prozessoren.

- Homogene Multicores
- Heterogene Multicores

Der Hauptunterschied dabei ist, dass bei den homogenen Multicores alle Kerne gleichartig sind. Dies hat den Vorteil, dass ein solcher Multicore einfacher zu fertigen ist, da man einfach ein Kerndesign dupliziert und geeignet miteinander verbindet. Auch für den Programmierer und den Scheduler des Betriebssystems ist es einfacher, solche Kerne zu verwalten, da man nicht darauf achten muss, auf welchem Kern ein Programm gerade läuft. Nachteilig wirkt sich allerdings aus, dass homogene Kerne nicht an spezielle Aufgaben angepasst sind und daher nicht immer mit optimaler Effizienz arbeiten können. Bei heterogenen Kernen ist es genau anders herum: Heterogene Kerne können für spezielle Aufgaben angepasste Kerne bereitstellen, die dann eine spezifische Aufgabe schnell und mit geringem Leistungsverbrauch erledigen können. Sowohl das Hardwaredesign als auch die Programmierung solcher Prozessoren können aber schwieriger zu handhaben sein. Die meisten auf dem Markt befindlichen Multicores sind immer noch homogen; heterogene Lösungen findet man hingegen oft im Embedded-Bereich.

## 1.6 Technologien zur Verbindung der Kerne

Bei wachsender Kernzahl ist es von zentraler Bedeutung, wie die verschiedenen Kerne mit den meist integrierten Speichercontrollern, anderen Kernen und den restlichen auf dem Prozessor integrierten Bausteinen verbunden werden. Dabei gibt es u.a. diese Architekturen:

**Bus** Eine Datenleitung, auf die alle angeschlossenen Teilnehmer (je einer pro Zeiteinheit) mit jedem anderen Teilnehmer Daten austauschen können.

**Punkt-zu-Punkt-Verbindung** Eine direkte Verbindung von zwei Bausteinen.

**Crossbar Switch** Ein Switch, der je einen Teilnehmer mit einem anderen direkt verbindet. Alle verbundenen Teilnehmer können parallel Daten übertragen, und die Verbindungen können nach einer Übertragung geändert werden.

**Packet Routing** Ein intelligentes Netzwerk, welches Datenpakete dynamisch und möglichst kostenminimal von der Quelle zum Ziel leitet.

Außerdem gibt es noch verschiedene Topologien, die zum Einsatz kommen:

- Ring
- Mesh

Die Ring-Topologie wird gerne mit dem Bus benutzt, daraus entsteht dann ein (oft bidirektionaler) Ringbus. Das Mesh hingegen wird meist mit der Punkt-zu-Punkt-Verbindung oder dem Packet-Routing benutzt und ist meistens als 2D-Grid konzipiert.

# Kapitel 2

## CPUs

Nach den einführenden Betrachtungen soll jetzt der Aufbau der CPUs näher betrachtet werden. Eine klassische CPU entspricht dem SISD-Modell, wobei es seit langer Zeit auch Erweiterungen in Richtung SIMD gibt. Den Anfang hat hierbei MMX von Intel gemacht. Mittlerweile gibt es eine große Menge von Befehlssatzerweiterungen, die die SIMD-Fähigkeiten von CPUs erweitern (SSE1,2,3,4 etc.). Es sind in Zukunft weitere SIMD-Erweiterungen geplant (z.B. AVX), um die Rechenleistung zu erhöhen.

Weiterhin wird eine CPU dadurch charakterisiert, dass die von-Neumann-Architektur beliebige Sprünge in den Instruktionen und den Daten erlaubt, welche zwar bei der Bearbeitung von Grafikdaten auf der GPU auch möglich sind, aber eher selten auftreten. Außerdem besitzen CPUs (im Gegensatz zu GPUs) einen von der Hardware verwalteten Cache, um den sich der Programmierer nicht selbst kümmern muss. Der Nachteil hierbei ist, dass der transparente Cache sehr viel Transistoren braucht. Bei aktuellen Intel-Modellen bilden durchaus mehr als ein Drittel aller Transistoren der CPU den Cache. Im folgenden werden einige aktuell erhältliche CPUs genauer beschrieben. Im Gegensatz zum Vortrag wird hier auch verstärkt auf den Desktopbereich eingegangen, um auch ein Bild davon zu erhalten, was für die Masse der Benutzer an Hardware erhältlich ist.

### 2.1 IBM PowerXCell 8i

Der PowerXCell 8i ist eine verbesserte Version der ursprünglichen Cell Broadband Engine, die für die Sony PS3 entwickelt wurde. Während die Cell Version in der Sony PS3 und auch die damals separat von IBM erhältlichen Modelle nur für Gleitkommazahlen einfacher Genauigkeit ausgelegt waren und daher doppelt genaue Gleitkommazahlen nur langsam berechnen konn-

ten, hat die 2007 aktualisierte CPU-Version native Unterstützung für doppelt genaue Gleitkommazahlen und kommt damit ungefähr auf die fünffache Leistung des Originalmodells in dieser Disziplin.

Der PowerXCell 8i ist auch eine der wenigen CPUs auf dem Markt, die heterogene Multicores besitzt. Als betriebssystemfähiger Steuerkern kommt dabei ein PowerPC Kern mit zwei Hardware-Threads zum Einsatz, der im Gegensatz zum normalen PowerPC Kern jedoch in-order arbeitet. Die anderen acht Kerne, SPUs genannt, sind die Kerne für die eigentlichen Berechnungen.[14]

## 2.2 Intel

Intel hat vor einiger Zeit die neuen Nehalem CPUs als Core i-Serie eingeführt. Mittlerweile basieren alle CPUs auf dem neuen Design, vom Atom im unteren Leistungssegment über die Core i3/5/7 CPUs im Desktopmarkt bis zu den Xeons im Servermarkt. Die Neuerungen gegenüber dem Vorgängermodell Core2 beinhalten:

**Quickpath Interconnect** Beim Core2 wurde der Speicher und andere CPUs per FSB (Front Side Bus) angesprochen. Der Chipsatz und andere CPUs werden jetzt über die Hochgeschwindigkeits-Punkt-zu-Punkt-Verbindung Quickpath Interconnect angebunden.

**Turbo Boost** Wenn die CPU kühl genug ist und die TDP nicht auslastet, dann takten sich die Kerne automatisch hoch. Je mehr Kerne dabei in einem tiefen Schlafmodus liegen, desto höher können sich die restlichen Kerne hochtakten.

**Hyperthreading** Hyperthreading wurde bei fast allen CPUs wieder eingeführt, nachdem die Technik seit dem Pentium4 nicht mehr zur Anwendung kam.

**Echter Multicore** Alle Kerne befinden sich auf einem Die, wogegen beim Core2 Quad Prozessor zwei Dual Cores auf der IO-Platine waren, die gegenseitig über den FSB kommunizieren mussten.

**Integrierter Speichercontroller** Der Dual Channel-Speichercontroller wurde vom Chipsatz in das CPU-Die integriert.

**GPU Kern** Das bisher im Chipsatz integrierte GPU-Die wurde auf die Prozessor-IO-Platine integriert, jedoch nur bei den Dual Cores und den Atom CPUs.

**L3 Cache** Es wurde ein L3 Cache eingeführt, der von allen Kernen gemeinsam benutzt werden kann.

Die Verbesserungen hatten eine deutliche Steigerung der pro-Kern Leistung der Core-i-Serie zur Folge. Ansonsten ist Intel mit dem direkten Konkurrenten AMD zumindest gleichgezogen, was die grundlegende Architektur angeht.

## 2.3 AMD

AMD hatte schon seit dem Athlon64 (interner Kernname K8) einen integrierten Speichercontroller und ein echtes Multicore-Design, also alle Cores auf einem Die vereint. Den FSB gibt es seit dem Athlon64 auch nicht mehr; seitdem benutzt AMD ebenfalls eine serielle Hochgeschwindigkeits-Verbindung namens Hypertransport. In der pro-Kern Leistung ist AMD leider im Vergleich mit Intel deutlich im Nachteil, da das Design des aktuellen Phenom II (interner Kernname K10) immer noch auf der Architektur seines Vorgängers Athlon64 basiert. Diese wurde zwar, so weit es ging, optimiert und verbessert, jedoch merkt man dem Kerndesign sein Alter mittlerweile an (der erste Vertreter mit K8-Kern kam 2003 heraus).

Einige nennenswerte Optimierungen gibt es dennoch beim K10-Kern:

**L3 Cache** Es wurde ein L3 Cache eingeführt, der von allen Kernen gemeinsam benutzt werden kann.

**Turbo Core** Beim aktuellen Phenom II X6 Hexacore gibt es eine ähnliche Technik wie Turbo Boost von Intel, jedoch funktioniert das Hochtakten der Kerne nur dann, wenn mindestens die Hälfte der Kerne schlafen.

**Verbesserter Speichercontroller** Der Dual Channel Controller wurde mit zwei Modi versehen: der ganged mode entspricht dem klassischen Dual Channel Speicherzugriff, also ein Thread kann mit 128Bit auf das RAM zugreifen. Neu ist der ungedunged mode, der parallele Threads beschleunigen soll. Dabei ist es zwei Threads gleichzeitig möglich, mit 64Bit auf das RAM zuzugreifen.

In der wichtigen pro-Kern Leistung wird AMD wohl erst mit der nächsten komplett neuen Architektur Bulldozer (K12-Kern, siehe unten bei den kommenden Projekten) wieder zu Intel aufholen können. Zumindest in einem Segment ist AMD deutlich besser aufgestellt als Intel: Die Preise für AMD CPUs sind im Moment konkurrenzlos günstig. Nach dem Kauf von ATI sieht es aber für AMD zumindest im GPU-Bereich wesentlich besser aus.

# Kapitel 3

## GPGPUs

GPGPUs sind eine Erweiterung der GPU-Architektur, welche ursprünglich nur für die Bearbeitung bzw. Generierung von 3D-Grafiken ausgelegt war. Grafikberechnungen sind meist gut parallelisierbar, weswegen GPUs schon von Anfang an auf SPMD ausgelegt waren. Auffallend beim Betrachten der Transistoren von GPUs ist, dass sie im Gegensatz zu CPUs praktisch keine Die-Fläche für Cache benutzen. Das liegt daran, dass GPUs zwar einen kleinen Cache haben, jedoch die komplette Cache-Verwaltung vom Programmierer erledigt werden muss. Daher kann sich eine GPU die Hardwareverwaltung für einen transparenten Cache sparen.

### 3.1 Nvidia

Nvidia hat vor kurzem eine komplett neue Architektur herausgebracht (Code-Name Fermi). Die erste Chipversion, die veröffentlicht wurde, war sogar nicht auf Spieler, sondern auf GPGPU-Anwendungen zugeschnitten. So hat Nvidia beispielsweise Speicherchips mit Fehlerkorrektur (ECC - Error correction code) auf die Karten installiert, um die bei dem hoch getakteten RAM manchmal auftretenden (und in Spielen in geringer Zahl vernachlässigbaren) Speicherfehler zu korrigieren. Außerdem lag der Speicherausbau bei maximal 6GB, was sehr viel für eine Grafikkarte ist.

Fermi ist die bisher komplexeste GPU, die je gebaut wurde. Das ist schon an den gut 3 Milliarden Transistoren ersichtlich. Das Projekt wurde sehr oft verzögert, der Produktstart immer wieder verschoben. Das lag daran, dass es interne Probleme gab, die ein teilweises Redesign des Chips notwendig machten. Weiters hatte Nvidia mit thermischen und Fertigungs-Problemen zu kämpfen. Eigentlich sollten ursprünglich die 512 Kerne höher takten als die jetzt verkaufte Version. Da jedoch sowohl die Ausbeute bei der Produkti-

on nicht hoch genug war und einige Streaming Multiprocessors defekt waren, als auch die Wärmeabgabe problematisch war, hat sich Nvidia dazu entschlossen, einige Streaming Multiprocessors zu deaktivieren und damit mehr funktionierende Chips aus der Produktion zu bekommen und die Verlustleistung zu begrenzen.

Es gibt mittlerweile einige neuere Versionen des Chips, die jedoch wieder eher auf Spiele optimiert sind. Eine Überarbeitung des hier vorgestellten Fermi-Chips soll evtl. noch im Dezember 2010 erscheinen.[10][9]

## 3.2 AMD

Auch AMD hat starke GPUs im Sortiment. Der aktuell schnellste Chip, der Radeon HD 5870, ist mit gut 2 Milliarden Transistoren nicht ganz so komplex wie Nvidia's Fermi. Die Entwicklung verlief bei AMD auch deutlich problemloser. Trotz allem hat AMD im Moment mit dem 5870 die Leistungskrone inne. Vor allem die Leistung bei einfacher Genauigkeit ist deutlich vor Nvidia's Fermi Chip.[24]

## 3.3 Intel

Intel wollte eine neue GPU auf Basis vieler kleiner einfacher in-order X86-Kerne mit Codenamen Larrabee anbieten. Leider ist das Produkt aufgrund der nicht konkurrenzfähigen Leistung zu Nvidia's und AMD's Karten nie auf den Markt gekommen. Larrabees Design geht auf den Forschungschip Polaris zurück, den Intel 2007 gebaut hatte. Intel hat aus der Entwicklung von Larrabee jedoch wichtige Erkenntnisse gewonnen. Daher haben sie zwei Nachfolgeprojekte ins Leben gerufen:

**Knights Ferry** Dieses Projekt entspricht im Grunde genommen Larrabee. Es ist eine Art Kombination einer CPU, die für Gleitkommaberechnungen eine separate Einsteckkarte mit Vektoreinheiten nutzt. Knights Ferry soll für Entwickler bereitstehen und die Grundlage für das Nachfolgeprojekt Knights Corner bilden.

**Knights Corner** Dies ist das eigentliche Projekt, das als Ziel hat, eine Karte mit 50 Kernen mit mehr als 500 GFLOPS Leistung und maximal 75W TDP zu entwickeln. Dabei sollen zur Softwareentwicklung dieselben Tools benutzt werden können, die bereits jetzt für Xeon-Prozessoren bereitstehen.[3][5]

# Kapitel 4

## Ausblick auf zukünftige Entwicklung

Das letzte Kapitel soll einen Einblick über die Änderungen in der nächsten Prozessor-Generation, sowie eine Prognose für die Zukunft geben.

### 4.1 APUs

AMD hat diese Bezeichnung eingeführt. Sie bezeichnet laut AMD nicht nur die Integration der GPU auf das CPU-Die, sondern auch eine Änderung an der CPU. Sie soll in Zukunft für Gleitkommaberechnungen die GPU benutzen, um daraus einen deutlichen Geschwindigkeitsvorteil zu ziehen.

#### 4.1.1 Intel SandyBridge

Die nächste Generation von Intel wird wieder einige Neuerungen im CPU-Bereich bringen. Dazu gehört ein verbesserter Turbo Boost, die Integration der GPU auf dem CPU-Die und eine AVX-Einheit. Die AVX-Einheit ist eine neue SIMD-Komponente, die Matrixberechnungen enorm beschleunigen soll. Bei SandyBridge gibt es zwei Besonderheiten zu beachten:

1. Obwohl die GPU in die CPU integriert wird, bleiben die Einheiten logisch getrennt. GPU/CPU werden lediglich mit einem Ringbus mit dem Speichercontroller verbunden. Daher bezeichnet AMD SandyBridge nicht als APU.
2. Die AVX-Einheit in SandyBridge wird den Fused-Multiply-Add-Befehl nicht enthalten, unterstützt also die AVX-Erweiterung nicht komplett.

Außerdem ist speziell zu Intel noch erwähnenswert, dass die Firma aktuell das deutlich überlegene Entwicklungsmodell besitzt. Intel fährt seit 2007 das so genannte Tick-Tock-Modell. Ein Tick bezeichnet eine Verkleinerung der Strukturbreiten der vorhergegangenen Architektur, und ein Tock eine neue Architektur. Intel hat sich selbst auferlegt, jedes Jahr abwechselnd einen Tick, dann einen Tock durchzuführen, und hält diese Regel seit der Einführung auch ein.[2]

### 4.1.2 AMD

AMD hat die ehrgeizigsten Pläne für die Zukunft. 2011 stehen zwei neue Architekturen an:

**Llano** soll die erste richtige APU werden. Eine GPU, leistungsmäßig auf dem Niveau einer Radeon HD 5350, wird in die CPU integriert und für Gleitkommaberechnungen der CPU genutzt. Die höhere Performanz der GPU soll einen Geschwindigkeitsschub bewirken.[11]

**Bulldozer** ist der Codename für die erste komplett neue Architektur seit dem K8-Kern. Diese APU soll in der pro-Kern Leistung wieder zu Intel aufschließen. Außerdem wird damit erstmals bei AMD eine Art von Hyperthreading eingeführt. Bulldozer wird auch die AVX-Befehlsweiterungen komplett unterstützen (im Gegensatz zu SandyBridge).[6][12]

Die Situation stellt sich so dar, dass AMD durch den Kauf der Firma ATI im GPU-Bereich mehr Know-How als Intel hat und daher in der GPU-Leistung vorne liegt. Bei der pro-Kern CPU-Leistung ist AMD jedoch momentan schlecht aufgestellt. Es bleibt abzuwarten, ob das Bulldozer-Design konkurrenzfähig sein wird. Falls ja, hätte man Intel wieder ein zumindest gleichwertiges Produkt entgegenzusetzen.

## 4.2 Langzeitprognose

Prognosen auf lange Zeit sind schwierig, da sich der Hardwaremarkt häufig und schnell ändert. Es deutet sich allerdings an, dass es mit der aktuellen Entwicklungsstrategie nicht mehr sehr lange weitergehen wird. Zu Pentium4-Zeiten gab es ja das Taktraten-Rennen, mittlerweile hat sich die Situation in eine Art Kernzahl-Wettrennen gewandelt.

Dabei vertrauen die Hardware-Hersteller hauptsächlich darauf, dass sich mit Verkleinerungen der Strukturbreiten die Wärmeabgabe der Chips weiter senken lässt, was jedoch nicht unendlich möglich ist. Die Breite der Leiterbahnen nähert sich bald wenigen Atomlagen; weniger als eine Atomlage ist

natürlich nicht möglich. Die Lithographie-Maschinen, die bei der Produktion von Halbleitern eingesetzt werden, stoßen auch an physikalische Grenzen, da die Erzeugung von extrem hochfrequentem Licht sehr schwierig ist. Nicht zuletzt die Chips selbst bekommen immer größere Probleme, da bei diesen extrem kleinen Abmessungen quantenmechanische Effekte auftreten und z.B. Leckströme verursachen. Gordon Moore hat selbst in einem Interview gesagt, dass sein berühmtes Gesetz tot sei und es nicht mehr sehr lange so weitergehen werde.[8]

Ein Trend hin zu APUs ist deutlich erkennbar, wobei die darin enthaltenen GPUs aufgrund der produzierten Verlustleistung nie die Leistungsfähigkeit einer dedizierten Grafikkarte erreichen werden. Daher werden für die nähere Zukunft APUs und GPUs nebeneinander existieren und sich evtl. auch gegenseitig ergänzen.[4]

In den nächsten 10 Jahren wird sich vielleicht nicht viel an der bisherigen Entwicklungsstrategie ändern, aber danach wird man irgendwann umdenken müssen und evtl. mehr in Richtung heterogene Multicores gehen, da diese spezielle Aufgaben wesentlich energieeffizienter erledigen können.

# Literaturverzeichnis

- [1] Andreas Stiller. Neue Hemispären. *c't*, pages 142–145, 21/2010.
- [2] Andreas Stiller. Intels sandige Brücken. *c't*, pages 18–20, 21/2010.
- [3] Computerbase. Intel Knights Ferry inklusive Chip untersucht, 2010. URL <http://www.computerbase.de/news/hardware/grafikkarten/2010/september/intel-knights-ferry-inklusive-chip-untersucht/>. [Ganzer Artikel].
- [4] Godfrey Cheng. Discrete GPUs: Dead, Mostly Dead, or Alive and Thriving?, 2010. URL <http://blogs.amd.com/play/2010/08/06/amd-discrete-graphics-gpus-fusion-apu/>. [Ganzer Artikel].
- [5] HPCWire. Compilers and More: Knights Ferry Versus Fermi, 2010. URL <http://www.hpcwire.com/features/Compilers-and-More-Knights-Ferry-v-Fermi-100051864.html>. [Ganzer Artikel].
- [6] ht4u.net. Vorhang auf für Bulldozer - AMD enthüllt neue Architektur, 2010. URL [http://ht4u.net/reviews/2010/amd\\_bulldozer\\_preview/](http://ht4u.net/reviews/2010/amd_bulldozer_preview/). [Ganzer Artikel].
- [7] IBM. Cell Broadband Engine Architecture and its first implementation, 2005. URL <http://www.ibm.com/developerworks/power/library/pa-cellperf/>. [Ganzer Artikel].
- [8] Manek Dubash. Moore's Law is dead, says Gordon Moore, 2005. URL <http://news.techworld.com/operating-systems/3477/moores-law-is-dead-says-gordon-moore/>. [Ganzer Artikel].
- [9] Martin Fischer, Andreas Stiller. Im Zeitalter des CUDAismus - High Performance Computing mit Nvidias Fermi. *c't*, pages 106–109, 19/2010.

- [10] Nvidia. Tesla C2050 / C2070 GPU Computing Processor, 2010. URL [http://www.nvidia.com/object/product\\_tesla\\_C2050\\_C2070\\_us.html](http://www.nvidia.com/object/product_tesla_C2050_C2070_us.html). [Ganzer Artikel].
- [11] Planet3DNow! AMD zeigt Llano APU auf der TFE 2010, 2010. URL <http://www.planet3dnw.de/cgi-bin/newspub/viewnews.cgi?id=1287474011>. [Ganzer Artikel].
- [12] tecchannel. AMD enthüllt CPUs Bulldozer und Bobcat, 2010. URL [http://www.tecchannel.de/server/prozessoren/2030475/amd\\_enthueellt\\_bulldozer\\_und\\_bobcat\\_cpu\\_fuer\\_server\\_pc\\_netbook/](http://www.tecchannel.de/server/prozessoren/2030475/amd_enthueellt_bulldozer_und_bobcat_cpu_fuer_server_pc_netbook/). [Seite 1, 2].
- [13] Tiler. TILE-Gx processor family, 2010. URL [http://www.tilera.com/products/processors/TILE-Gx\\_Family](http://www.tilera.com/products/processors/TILE-Gx_Family). [Ganzer Artikel].
- [14] Wikipedia. Cell (microprocessor) — Wikipedia, The Free Encyclopedia, 2010. URL [http://en.wikipedia.org/w/index.php?title=Cell\\_\(microprocessor\)&oldid=394140557](http://en.wikipedia.org/w/index.php?title=Cell_(microprocessor)&oldid=394140557). [1 History - 3 Architecture].
- [15] Wikipedia. Mehrkernprozessor — Wikipedia, Die freie Enzyklopädie, 2010. URL <http://de.wikipedia.org/w/index.php?title=Mehrkernprozessor&oldid=80162300>. [4 Liste einiger Mehrkernprozessoren].
- [16] Wikipedia. Multi-core processor — Wikipedia, The Free Encyclopedia, 2010. URL [http://en.wikipedia.org/w/index.php?title=Multi-core\\_processor&oldid=393547391](http://en.wikipedia.org/w/index.php?title=Multi-core_processor&oldid=393547391). [6 Hardware examples].
- [17] Wikipedia. Hardwareseitiges Multithreading — Wikipedia, Die freie Enzyklopädie, 2010. URL [http://de.wikipedia.org/w/index.php?title=Hardwareseitiges\\_Multithreading&oldid=80992218](http://de.wikipedia.org/w/index.php?title=Hardwareseitiges_Multithreading&oldid=80992218). [Ganzer Artikel].
- [18] Wikipedia. Multithreading (computer architecture) — Wikipedia, The Free Encyclopedia, 2010. URL [http://en.wikipedia.org/w/index.php?title=Multithreading\\_\(computer\\_architecture\)&oldid=394172768](http://en.wikipedia.org/w/index.php?title=Multithreading_(computer_architecture)&oldid=394172768). [2.3 Simultaneous multi-threading].
- [19] Wikipedia. Out-of-order execution — Wikipedia, The Free Encyclopedia, 2010. URL [http://en.wikipedia.org/w/index.php?title=Out-of-order\\_execution&oldid=375198633](http://en.wikipedia.org/w/index.php?title=Out-of-order_execution&oldid=375198633). [Ganzer Artikel].

- [20] Wikipedia. Opteron — Wikipedia, The Free Encyclopedia, 2010. URL <http://en.wikipedia.org/w/index.php?title=Opteron&oldid=393933632>. [2.5 Opteron (45 nm SOI), Dodeca-core – Magny-Cours MCM (6164-6176)].
- [21] Wikipedia. Pipeline (Prozessor) — Wikipedia, Die freie Enzyklopädie, 2010. URL [http://de.wikipedia.org/w/index.php?title=Pipeline\\_\(Prozessor\)&oldid=75776828](http://de.wikipedia.org/w/index.php?title=Pipeline_(Prozessor)&oldid=75776828). [Ganzer Artikel].
- [22] Wikipedia. Teraflops Research Chip — Wikipedia, The Free Encyclopedia, 2010. URL [http://en.wikipedia.org/w/index.php?title=Teraflops\\_Research\\_Chip&oldid=361617651](http://en.wikipedia.org/w/index.php?title=Teraflops_Research_Chip&oldid=361617651). [Ganzer Artikel].
- [23] Wikipedia. POWER7 — Wikipedia, The Free Encyclopedia, 2010. URL <http://en.wikipedia.org/w/index.php?title=POWER7&oldid=392834703>. [2 Specifications].
- [24] Wikipedia. Evergreen (GPU family) — Wikipedia, The Free Encyclopedia, 2010. URL [http://en.wikipedia.org/w/index.php?title=Evergreen\\_\(GPU\\_family\)&oldid=394298596](http://en.wikipedia.org/w/index.php?title=Evergreen_(GPU_family)&oldid=394298596). [3.2 Radeon HD 5800].
- [25] Wikipedia. Simultaneous Multithreading — Wikipedia, Die freie Enzyklopädie, 2010. URL [http://de.wikipedia.org/w/index.php?title=Simultaneous\\_Multithreading&oldid=75359953](http://de.wikipedia.org/w/index.php?title=Simultaneous_Multithreading&oldid=75359953). [Ganzer Artikel].
- [26] Wikipedia. UltraSPARC T2 — Wikipedia, The Free Encyclopedia, 2010. URL [http://en.wikipedia.org/w/index.php?title=UltraSPARC\\_T2&oldid=385423390](http://en.wikipedia.org/w/index.php?title=UltraSPARC_T2&oldid=385423390). [Ganzer Artikel].
- [27] Wikipedia. Xeon — Wikipedia, The Free Encyclopedia, 2010. URL <http://en.wikipedia.org/w/index.php?title=Xeon&oldid=393403498>. [6.7 6500/7500-series Beckton].