

Die IA-64

Der Parallelrechner auf einem Chip

IA-64, Epic und Itanium

Intels Antwort auf den langjährigen Zwist zwischen RISC und CISC heißt EPIC: Explicit Parallel Instruction Computing. Die Hauptaufgabe moderner multiskalarer Prozessoren ist nämlich das Verteilen der anstehenden Aufgaben auf viele parallel arbeitende Einheiten. Statt dass eine sehr komplexe Hardware jedes mal auf Neue die aufeinander folgenden Befehle analysiert und versucht diese parallel auszuführen, ist das bei EPIC die Arbeit des Compilers. Dieser bemüht sich den Code in möglichst optimal parallel ausführbare Häppchen zu zerlegen. Bei IA-64 besteht so ein Häppchen aus einem Bündel von drei Befehlen zu je 41 Bit + das sogenannte Template, das darüber Auskunft gibt, was das für Befehle sind und wo die Parallelität endet. Es ist auch möglich, dass das nächste Bündel parallel zum ersten ausführbar ist, und theoretisch auch das übernächste und so weiter. Es sind allerdings nicht alle Befehlsarten in den Bündeln beliebig mischbar. Z.B kann jedes Bündel nur einen einzigen Gleitkommabefehl enthalten. Anders als bei den RISC-Prozessoren kann IA-64 auch direkt ein 64-Bit-Wort an ein Register übergeben. Das würde dann zwei Befehle belegen. Die Triplets in den Bündeln mit ihren Befehlsartenkürzeln M (Memory), I (integer), F (Floating Point Unit), B (Branch) oder LX (64 Bit Immediate) reihen sich zu einem Programm zusammen zum B.:MII-BBB-MFI

42,66-Bit-Prozessor

Von der Instruktionseite her ist IA-64 gar nicht 64-bittig. Auch die Adressenbreite ist nicht zwangsläufig 64-bittig, sondern beim jetzt vorgestellten Itanium physisch „nur“ 44-bittig, (für 16 TByte Speicher; entspricht etwa 64 000 Standard-PCs vom Speichervolumen mit 256 Mbyte) und virtuell 54-bittig. Vor allem ist es der große Adressraum, der die 64-Bitter auszeichnet. Die oberen drei Bits der virtuellen Adresse dienen als Kennung für verschiedene Adressraumtypen. Damit lassen sich zum Beispiel Betriebssystem Kernel auf User-Bereich festlegen. Andere Räume sind für „Shared Memory“ sinnvoll oder für solche, die nur große Pages beinhalten. IA-64 unterstützt alle Page-Größen in Zweier-Potenzen zwischen 4Kbyte und 256 Mbyte. Diese Pages dienen der Umrechnung zwischen den virtuellen und den physischen Adressen, so wie es alle modernen Betriebssysteme erfordern. Damit diese Umrechnung nicht zeitaufwendig bei jedem Speicherzugriff nötig ist, enthalten die Prozessoren Zwischenspeicher(Translation Lookaside Buffer, TLB).Jeder Prozess hat dank dieser Umrechnung seinen eigenen Adressraum und Page-Verzeichnisse. IA-64 hat hierfür „datenbankähnliche“ Strukturen mit Hashes und spezielle Hardware vorgesehen, um das Durchstöbern der Einträge, flott zu bewältigen. Außerdem kann das Betriebssystem direkt, mit speziellen Befehlen versteht sich auf das TLB einwirken.

Register, Register-Management

IA-64 bietet mit 128 Allzweck-Registern(r0...r127) 16-mal mehr Register als IA-32, die zudem allen 64-Bittig sind (+das Not a Thing -Bit ,NaT). Diese 128 Register teilen sich in zwei Bereiche: die erste 32 sind statisch und die restlichen rotieren, werden also dynamisch verwaltet. D.h. der Compiler gibt bei einer Unteroutine an, wie viele Register er für Ein- und Ausgabe und lokale Berechnungen benötigt. (vermeiden übermäßige push und pop wie beim IA-32 nötig).

IA-64 stellt ebenso 128 Gleitkommaregister zur Verfügung, jene sollen den herausragenden Schwachpunkt der Gleitkommaberechnung der alten IA-32 Architektur den gar aus machen.

Auch sie teilen sich auf in 32 statische und 96 dynamische, mit eben diversen Vorteilen und Rotationsmöglichkeiten. Die FP-Register sind mit 82 Bit sogar etwas breiter als bei IA-32 (80 Bit). Die höhere 82-Bit-Genauigkeit hat allerdings einen praktischen Hintergrund, und zwar erfordert die Abspeicherung der Zwischenwerte die bei der Berechnung von Division, Wurzel, Sinus, Tangens, Exponentialfunktion und Logarithmus anfallen die 2 zusätzlichen Bits. Übrigens stellt diese Funktionen der Compiler mittels Softwarebibliothek zur Verfügung, nicht wie bisher die Instruction Set Architecture.

Neben den GP- und FP-Registersätzen unterstützt IA-64 noch eine Vielzahl von Spezialregister:

128 Applikations-, 8 Sprung-, 8 COUID- und diverse Performance – Monitor-Register für die Performance-Counter.

Der jetzt offiziell vorgestellte Itanium mit dem Codenamen Merced stellt die in 25 Millionen Transistoren „gegossene“, Implementierung der neuen 65-Bit Architektur dar, mit vier Integer-, vier Multimedia- und zwei Gleitkommaeinheiten. Dazu kommen noch zwei weitere Gleitkommaeinheiten mit Single-Precision, drei Sprung, und zwei Lade-/Speichereinheiten. Die 17 Funktionseinheiten (2FPU, zwei Memory, zwei Integer und drei Branches) werden über 9 Ports mit den nötigen Bits versorgt. Diese warten in einem Buffer (Caches).

Beim IA-64 aufgeteilt in L1-Caches 16 KB, für Instruktionen und Daten, L2-Caches 96KB, beide on board, und der L3-Cach sitzt extern auf einem Modul mit der Größe 2 oder 4 MByte.

EPIC (die nächste Generation)

EPIC (Explizit parallel instruction computing) ist eine Weiterentwicklung von bereits vorhandenen Techniken (CISC, RISC, VLIW, OOO, ...). Um diese explizite Parallelität zu unterstützen ist eine Parallelisierung des Maschinencodes notwendig (ILP = instruction level parallelism). Bei dieser Technik fällt die Parallelisierungslast nicht mehr auf die Hardware, sondern auf den Compiler.

Weiters braucht man für eine ILP – Architektur zusätzliche Techniken zur Spekultativen Ausführung, Datenvorhersage und Sprungvorhersage.

Sehr wichtig für ILP ist sehr großer Registersatz, mehrere Funktionseinheiten (Rechenwerke) und ein supersklares Leitwerk.

Instruction groups

Die IA-64 erlaubt es Instruktionen in Bündel, die sogenannten Instruktion groups, zu gruppieren die gleichzeitig ausgeführt werden können. Dies hat den Vorteil dass die Hardware keine Registerabhängigkeiten mehr prüfen muss (dies wird vom Compiler gemacht).

Die Sprungarchitektur

Die IA-64 besitzt 3 verschiedene Typen von Sprüngen: die Offset Sprünge (21Bit), indirekte Sprünge via 8 Branch Registern und spezielle Sprungbehandlungen für Schleifen und Multi-Way-Sprünge (mit mehreren Sprungzielen). Weiters unterstützt die IA-64 sogenannte Hints. Das sind Vorhersagen des Compilers welchen Weg das Programm, bei der Verzweigung einschlagen wird.

Das Predication

Bei einem Vergleich wird das Ergebnis auf zwei Prädikate ausgegeben. Jedes Prädikat wird nun mit einem Sprungziel verknüpft und aufbauend auf diese Prädikate wird nun die Kondition ausgeführt.

Das Sprungziel (bestehend aus evt. mehreren Anweisungen) kann durch die Technik der Basic-Blocks schon im Voraus ausgeführt werden (spekulative Ausführung). Weiteres können

der „then“- und „else“-Zweig zeitgleich ausgeführt werden. So werden Zeitverluste bei Flaschevorhersagen eliminiert.

Parallel Compares

Die Technik der parallelen Vergleiche erlaubt es Konditionen zeitgleich auszuführen und ihre Prädikate mit logischen Funktionen (z.B. or oder and) zu Verknüpfen. So können verschachtelte Konditionen (wie $(a==1) \text{ or } (b==1)$) in einem Schritt berechnet werden.

Speculation

Control speculation

Die Control speculation bewegt Ladebefehle vor einen möglichen Sprung. Tritt beim Laden ein Fehler auf so wird das im NatBit vermerkt. An der Stelle, an der eigentlich der Ladebefehl stehen müsste, wird eine Checkroutine eingefügt, die das NatBit kontrolliert und im Fehlerfall eine Recovery-Routine, die den Ladebefehl neu ausführt.

Auch das spekulative Benützen der spekulativ geladenen Daten ist möglich. Dabei wird mit dem NatBit gleich vorgegangen wie oben.

Data speculation

Die data speculation bewegt Ladebefehle vor einen möglichen Überschreibebefehl. Die geladen Daten werden in der ALAT (Advanced load address table) vermerkt. Werden die Daten überschrieben wird der Eintrag aus der ALAT entfernt. Der Eintrag kann auch durch einen speziellen Befehl aus der ALAT entfernt werden. An der Stelle wo der eigentliche Ladebefehl stehen müsste, wird eine Checkroutine eingefügt, die kontrolliert ob in der ALAT die Daten vermerkt sind. Ist nun kein Eintrag vorhanden, so müssen die Daten neu geladen werden. Auch hier sind spekulative Benütungen erlaubt.

Die Daten- und Kontrollspekulation kann auch kombiniert werden.

Software Pipelining

Die Technik des Softwarepipelining ermöglicht es Schleifendurchgänge nebeneinander auszuführen. Im Konkreten sieht das so aus, dass während im ersten Schleifendurchlauf die Berechnung durchgeführt wird, schon der zweite Schleifendurchgang initialisiert wird. Das hat für bestimmte Anwendungen einen sehr großen Geschwindigkeitsvorteil, da z.B. statt zwei, vier Durchgänge ausgeführt werden können. Diese Technik benötigt aber einen ziemlich großen Hardwareaufwand und verwendet die oben erklärten Techniken des Predication und Spekulation.

Der Compiler

Nur neune auf den Prozessor optimierte Compiler sind in der Lage den vollen Funktionsumfang der IA-64 auszuschöpfen. Die INTEL und die HP arbeiten an Compiler für C/C++ und Fortran, die teilweise schon voll funktionsfähig sind.

Ein interessanter Aspekt der Compiler ist, dass nicht nur der Compiler der Hardware Tipps zur Ausführung (wie z.B. Hints) gibt, sonder auch die Hardware dem Compiler während der Compile-Zeit.

Kompatibilität zur IA-32

Benutzer von IA-32 Systemen verlangen von der IA-64 eine Abwärtskompatibilität. Für Programme die in höheren Programmiersprachen geschrieben sind, ist diese leicht durch

Rekompilation zu realisieren. Für Programme die auf Systemressourcen direkt zugreifen (wie z.B. OS, oder Treiber für Peripherie) ist die Kompatibilität nicht so leicht zu gewährleisten. Es gibt aber bereits einige Betriebssysteme für die IA-64, wie z.B. Win64 2000 oder HP-UX und das Trillian Project für Linux Anwender.

Für die jetzt noch verbleibenden IA-32 Programme bietet die IA-64 eine Hardware namens Dynamic-Translation die die IA-32 Anweisungen zur Laufzeit auf IA-64 Anweisungen übersetzt. Man hat kaum Leistungseinbusen bei der Ausführung von IA-32 Code, da der Leistungsverlust beim übersetzten der Befehle durch den großen Funktionsumfang der IA-64 ausgeglichen wird.

Der Intel Itanium

Der Intel Itanium ist der erste Prozessor der aufbauend auf der IA-64 Architektur realisiert wurde. Als Plattform für das System dient der Intel 460GX Chipsatz. Als Betriebssystem wurden Linux und Win2000 64 bereits auf wirklicher Hardware getestet.

Der Prozessor besteht aus 230Mio. Transistoren von denen sich 25Mio auf dem Prozessor selbst und 295Mio. auf dem L3 Cache befinden.

Kommentare

- IA-64 repräsentiert die wichtigste Entwicklung seit dem 80386er.
- IA-64 bewegt sich jenseits der traditionellen Architekturen
 - Compiler, Hardware Unterstützung, Massive Ressourcen
- IA-64 besitzt Tools um moderne high-end Anwendungen zu unterstützen
 - E-business
 - Technical computing
- IA-64 ist eine erweiterbare Architektur (MacKindley, Madison, Deerfield sind bereits in Entwicklung)
- Die IA-64 wird Nachfolger des Intel Xeon und demnach vor allem in Server- und Workstationbereich eingesetzt werden
- Also zunächst für die höchsten Leistungsklassen und nicht für den Privatanwender
- IA-64 ist größer als herkömmliche Prozessoren
- IA-64 hat einen größeren Leistungsverbrauch
- Intel genießt momentan recht viel Kritik an dieser Neuentwicklung

Links

IA-64

- <http://developer.intel.com>
- <http://developer.intel.com/design/ia64>
- <http://developer.intel.com/technology/itj/q41999.htm>

Trillian Project

- <http://www.linuxia64.org>