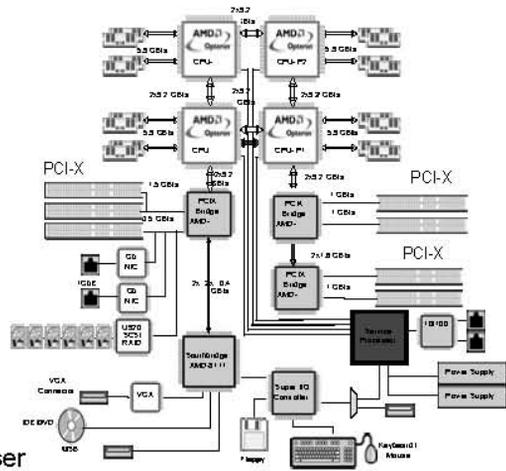


Wo ist der Flaschenhals?

- Memory: 5 GBps
- HyperTransport: 3 GBps
- Southbridge: 1 GBps
- PCI: 100. . .700 MBps
- HBA: 200. . .500 MBps
- FC: 200 MBps
- Device: 60...30 MBps
- Damit notwendig:  
2 PCI-X für HDs  
2 PCI-X für Netzwerk  
aber, PCI Express ist besser



Leistungssteigerung von Rechnern

Verarbeitungsleistung = Datendurchsatz \* Verarbeitungsgeschwindigkeit

- **Schnellere Befehlsabarbeitung**  
→ höhere Taktfrequenzen; Problem: Verlustleistung; Hochtechnologien
- **Vergrößerung der Speicherbandbreite**  
→ größere Datenbreiten; Dual Bus Archit.; höhere Datenraten; Hyper Transport; Cache vergrößern; zusätzliche Cache-Levels;
- **Vergrößerung der Verarbeitungsbreite**  
→ größere Datenbreite: 8 Bit → 16 Bit → 32 Bit → 64 Bit → ??
- **Architekturveränderungen**  
→ von CISC zu RISC; Pipelining; Out of Order; Super Skalare Architektur; Speculative Execution; Predication; Streaming SIMD Extension (SSE2,3); Spezialisierung (Abkehr vom Universalprozessor, applikationsspezifische µP)
- **Parallelverarbeitung**  
→ mehr Register; Hyper Threading; EPIC (Explicitly Parallel Instruction Computing); Dual/Multi Core; Multiprozessorsysteme (Parallelrechner)

Leistungskennzahlen von Rechnern

- Taktfrequenz
  - Wie viele elementare Operationen (nicht Maschinenbefehle!) können pro Sekunde ausgeführt werden
- (M)IPS ((Million) Instructions per Second)
  - Maschinenbefehle pro Sekunde
  - Komplexität der Befehle? Siehe CISC/RISC
- CPI (Clocks per Instruction)
  - Anzahl der Takte pro Maschinenbefehl
- Flops (Floating Point Operations per Second)
  - Aussagekräftig vor allem für wissenschaftliche, graphische, und andere sehr rechenaufwändige Anwendungen mit Gleitkommazahlen
  - Sind das Maß aller Dinge
- Andere Leistungsaspekte für Anwendungsleistung, z.B.:
  - Datendurchsatz für Ein-/ Ausgabe
  - Speicherzugriffe



Benchmarks

- Exakt vorgegebene und standardisierte Berechnungen bzw. Tätigkeiten mit Standard-Anwendersoftware
- Teilweise auf Anwendungsdomänen zugeschnittene Testverfahren (z.B. Office, HomeUser, Grafik, Web-Server, etc.)
- Synthetische Benchmarks, z.B.:
  - **Linpack, Whetstone, Dhrystone**
- Bedeutende Benchmark-Organisationen:
  - **SPEC** (Standard Performance Evaluation Corporation)  
SPEC CPU2000, SPEC web99, SPEC ViewPerf, SPEC JBB (Java Business Benchmark)
  - **BAPCo** (Business Applications Performance Corporation)  
PCMark2002, Mobile Mark2002, SysMark2002, 3DMark2001, WebMark2001



CPI-Wert

Unter dem **CPI-Wert** (Cycles Per Instruction) versteht man die Anzahl interner Bus-Zyklen (Takte) pro Maschinenbefehl.

Der CPI-Wert eines Programms ( $\overline{CPI}$ ) ergibt sich aus der Summe der Produkte von CPI-Werten gleich schneller Befehle und der relativen Häufigkeit  $h_i$  ihres Auftretens.

$$\overline{CPI} = \sum_{i=1}^n CPI_i \cdot h_i$$

- CISC-Maschinen:  $\overline{CPI} > 1 \dots 2$
- RISC-Maschinen:  $\overline{CPI} < 1$

Programmlaufzeit = Dauer eines Buszyklus \* Anzahl der auszuführenden Befehle \* CPI-Wert des Programms



MIPS – Million Instructions per Second

Die Bewertungszahl **MIPS** wurde in den 70er Jahren eingeführt und ist eine Maßzahl für die Anzahl von Maschinenbefehlen, die eine CPU pro Sekunde ausführt.

Als Vergleichsgröße für unterschiedliche Prozessoren konnten sie sich allerdings nicht durchsetzen, da unterschiedliche Architekturen zum Teil erheblich divergierende Befehlssätze haben.

Somit war kein objektiver Vergleich möglich.

MIPS wurde scherzhafter Weise auch als "Misleading Information to Promote Sales" oder "Meaningless Indication/Information about Processor Speed" bezeichnet.

CPU	Jahr	MIPS
Intel 8080 (2 MHz)	1974	0,6 MIPS
Motorola 68000 (8 MHz)	1979	1 MIPS
Intel 486DX (66 MHz)	1992	54 MIPS
ARM10 (300 MHz)	1998	400 MIPS
Athlon 64 (2,8 GHz)	2005	8400 MIPS



FLOPS - floating-point operations per second

**FLOPS** ist eine Maßzahl zur Geschwindigkeitsmessung von wissenschaftlich genutzten EDV-Systemen bzw. Prozessoren und wurden ebenfalls in den 70er Jahren eingeführt.

Man ermittelt dabei die Anzahl von Gleitkommazahloperationen (Addition und Multiplikation) die in einer Sekunde ausgeführt werden.

Die Anzahl Gleitkommazahloperationen ist im Gegensatz zu MIPS nicht zwingend proportional zur Taktfrequenz.

Unterschiedliche Architekturen/Prozessoren haben unterschiedliche Befehlsängen bzw. Implementierungen für Gleitkommazahloperationen. Vektorprozessoren führen beispielsweise in einem Takt mehrere tausend Operationen aus.

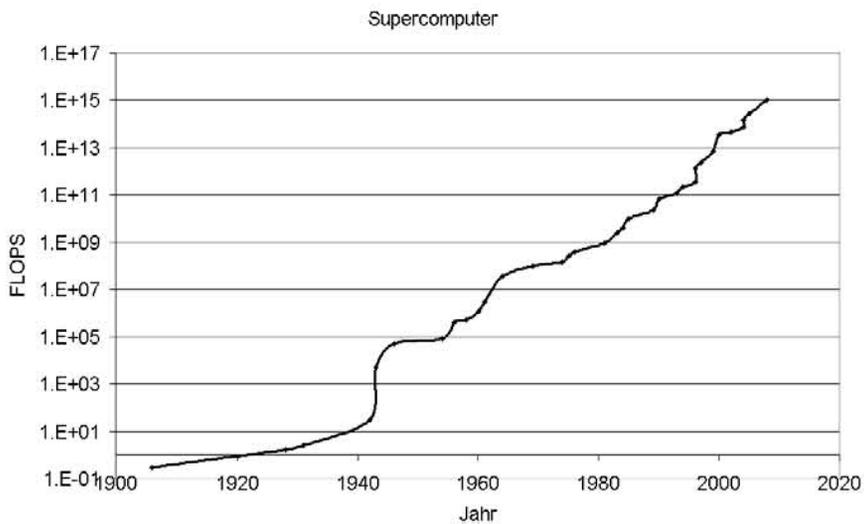
**Leistung heutiger Computer:**

- Dual Core PC : **10 ... 20 GFLOPS**
- Supercomputer: **> 500 TFLPOS**

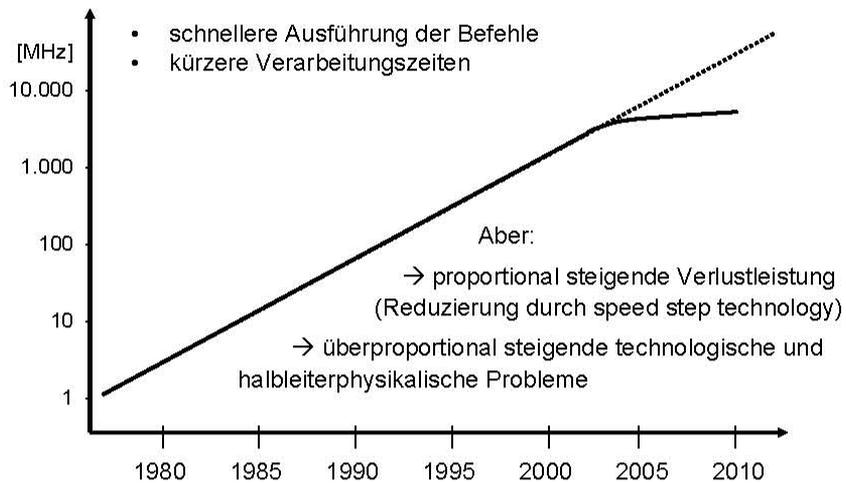


Die Leistung von Supercomputern

<http://de.wikipedia.org/wiki/Supercomputer>

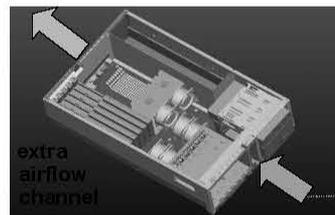


Erhöhung der Prozessortaktfrequenzen



Problem: Verlustleistung

- Jede neue Prozessorgeneration bringt mehr Performance und mehr Verlustleistung.
- Wird die Wärme nicht ausreichend abgeführt, sinkt die Zuverlässigkeit des Systems.
- Die Wärmeabfuhr spielt eine immer größere Rolle bei der Konzipierung von Rechnersystemen
  - Aufbau, Layout des Motherboards (siehe BTX)
  - Kühlungskonzepte, Gestaltung des Luftstromes, Flüssigkeitskühlung
  - Qualität und Zuverlässigkeit des Systems (MTBF)
  - Serverdichte in einem Rack,
  - Leistung pro m<sup>3</sup>
  - Gutes Power Management ist besser, als viele Lüfter



Von Neumannscher Flaschenhals

Leistung eines Rechnersystems hängt nur z.T. von der Prozessorleistung ab.

Die Schnittstelle zwischen Prozessor und Speicher wird als "von Neumannscher Flaschenhals" (engl. bottleneck) bezeichnet.

Problem → Daten werden vom Prozessor schneller verarbeitet als sie aus dem Speicher gelesen oder in den Speicher geschrieben werden können.

Geschwindigkeit der Prozessoren ist stärker gestiegen, als die peripherer Komponenten

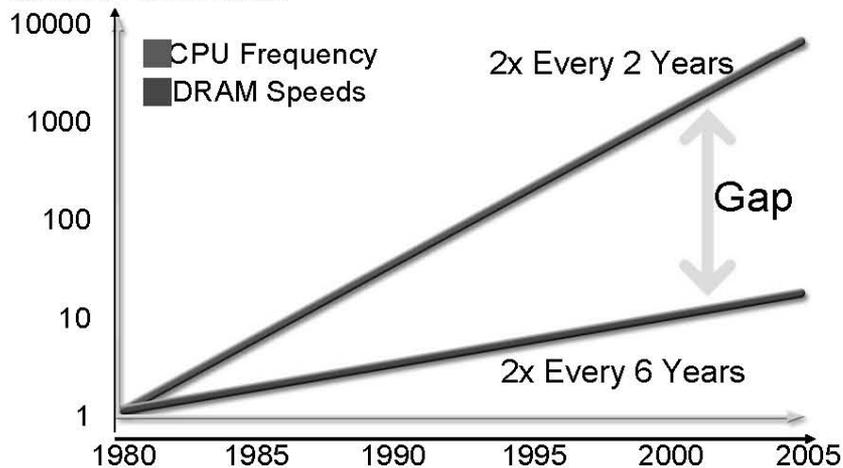
JAHR	CPU-Takt	Speichergröße	Speicherzugriff
1975	500 ns	4 Kbit	300 ns
2000	1... 1,5 ns	256 Mbit	30... 50 ns

CPU muss bei jedem Speicherzugriff bis zur Datenbereitstellung warten (Leistungseinbuße infolge eingefügter Waitstates)



Von Neumannscher Flaschenhals

Relative Performance



Von 32 zu 64 Bit Systemen

64 Bit Systeme erfordern neue Computertechnologien, die nicht mehr mit konventionellen 32 Bit Systemen realisierbar sind (z.B. würde Paging viel zu lange Rechenzeiten erfordern).

- 32 Bit Prozessoren → 4 GB adressierbarer Speicher
  - $2^{32}$  → 4.294.967.296 Byte (kann nicht komplett von Windows genutzt werden)
- 64 Bit Prozessoren → 8 TB adressierbarer Speicher
  - theoretisch:  $2^{64}$  → 18.446.744.073.709.551.616 Byte = 16 EByte → mehr, als ausreichend für Betriebssystem und Anwendungen
  - Für volle Unterstützung sind 64-Bit Treiber und 64 Bit Betriebssysteme erforderlich
- Vorteile der 64-Bit Verarbeitung
  - Doppelte Datenmenge pro Befehl, doppelte Verarbeitungsbreite
  - Verbesserte Multimedia-Verarbeitung (Fast Encoding/Decoding)
  - Schnelle rechenintensive Anwendungen (Graphics, Encryption, Research)
  - Aber, mehr Speicherbedarf für Zeiger zur Speicherverwaltung



Von 32 zu 64 Bit Systemen



- Völlig neue Architektur (Itanium)
- EPIC (Explicitly Parallel Instruction Computing)
- Predication
- Inkompatible mit 32 Bit Software
- Für Business-Anwendungen konzipiert
- EM64T – kompatible mit der AMD-Erweiterung



- Erweiterung der existierenden IA32 Architektur
- Existierende 32 Bit-Software läuft optimal
- Für Home/Small Office konzipiert



**Computertechnik** **Leistungssteigerung**

**Verbesserung der Prozessor-Architektur**

T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>	T <sub>7</sub>	T <sub>8</sub>	T <sub>9</sub>	T <sub>10</sub>
IF	ID	MA	Ex	WB	IF	ID	MA	Ex	WB

Befehl 1                      Befehl 2

Weg von der klassischen SISD-Struktur ...

... (Befehle werden einzeln und nur nacheinander ausgeführt.)

... hin zu Pipelining

1	IF	ID	MA	Ex	WB	IF	ID	MA	Ex	WB
2	IF	ID	MA	Ex	WB	IF	ID	MA	Ex	WB
3	IF	ID	MA	Ex	WB	IF	ID	MA	Ex	WB
4	IF	ID	MA	Ex	WB	IF	ID	MA	Ex	WB
5	IF	ID	MA	Ex	WB	IF	ID	MA	Ex	WB

Befehl →

... Ausführung in mehreren ... (parallelen) Stufen.

Schon den nächsten ... Befehl (Daten) holen, während der aktuelle ... gerade ausgeführt wird.

CPI < 1 möglich.

Prof. Dr.-Ing. W. Schmalwasser
Hochschule Mittweida - Fachbereich: IT & ET

**Computertechnik** **Leistungssteigerung**

**Verschiedene Pipeline-Konzepte**



- **Viele einfache Pipeline-Stufen**
  - Höhere Taktfrequenzen
  - Performanceverlust bei falscher Sprungvorhersage
  - Höhere Verlustleistung



- **Wenige, aber tiefe Pipeline-Stufen**
  - Während einer Taktperiode wird mehr getan (ausgeführt)
  - Keine hohen Taktfrequenzen erforderlich
  - Geringere Leistungsaufnahme

Prof. Dr.-Ing. W. Schmalwasser
Hochschule Mittweida - Fachbereich: IT & ET

**Computertechnik** **Leistungssteigerung**

**Verbesserung der Prozessor-Architektur**

T<sub>1</sub> T<sub>2</sub> T<sub>3</sub> T<sub>4</sub> T<sub>5</sub> T<sub>6</sub> T<sub>7</sub> T<sub>8</sub> T<sub>9</sub> T<sub>10</sub>

IF ID MA Ex WB IF ID MA Ex WB

-----

**Befehl 1**                      **Befehl 2**

Weg von der klassischen SISD-Struktur ...

... (Befehle werden einzeln und nur nacheinander ausgeführt.)

... hin zur superskalaren Architektur (SIMD-Struktur)

Mehrfach vorhandene Ausführungseinheiten - parallele Ausführung mehrerer Befehle (in einer Pipeline-Stufe)

Prof. Dr.-Ing. W. Schmatwusser Hochschule Mittweida - Fachbereich: IT & ET

**Computertechnik** **Leistungssteigerung**

**Verbesserung der Prozessor-Architektur**

T<sub>1</sub> T<sub>2</sub> T<sub>3</sub> T<sub>4</sub> T<sub>5</sub> T<sub>6</sub> T<sub>7</sub> T<sub>8</sub> T<sub>9</sub> T<sub>10</sub>

IF ID MA Ex WB IF ID MA Ex WB

-----

**Befehl 1**                      **Befehl 2**

Weg von der klassischen SISD-Struktur ...

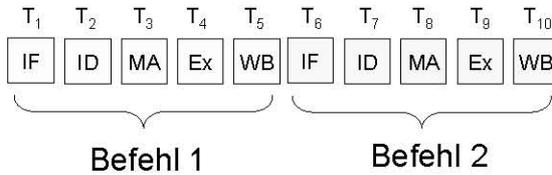
... (Befehle werden einzeln und nur nacheinander ausgeführt.)

... hin zur superskalaren Architektur (SIMD-Struktur)

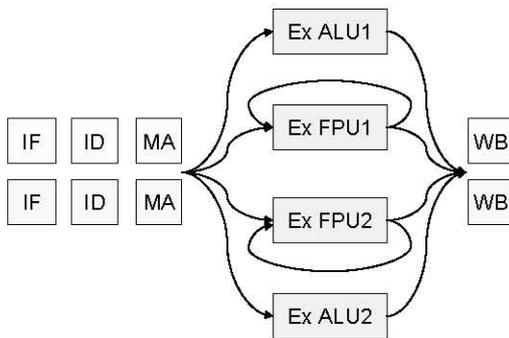
Mehrfach vorhandene Ausführungseinheiten - parallele Ausführung mehrerer Befehle (in einer Pipeline-Stufe)

Prof. Dr.-Ing. W. Schmatwusser Ex ALU2 Hochschule Mittweida - Fachbereich: IT & ET

Verbesserung der Prozessor-Architektur



Weg von der klassischen SISD-Struktur ...  
 ... (Befehle werden einzeln und nur nacheinander ausgeführt.)

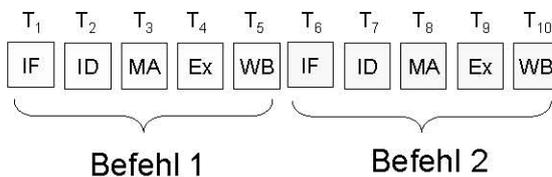


... hin zu Hyper-Threading (Befehlsebenenparallelität)

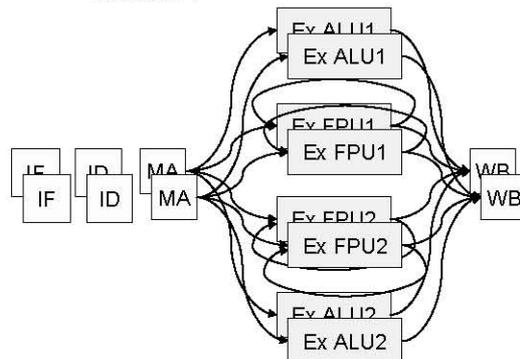
Abarbeitung von Threads in den ungenutzten mehrfach vorhandenen Komponenten eines Prozessors. Die Performance wird verbessert, aber nicht verdoppelt.



Verbesserung der Prozessor-Architektur



Weg von der klassischen SISD-Struktur ...  
 ... (Befehle werden einzeln und nur nacheinander ausgeführt.)

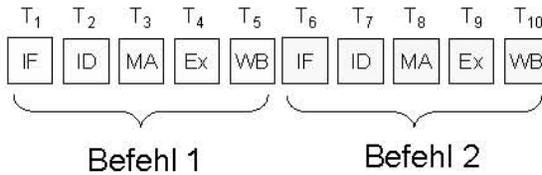


... hin zu Dual-Core

Zwei Prozessorkerne auf einem Chip - Zwei Threads laufen simultan und bringen fast die doppelte Performance.

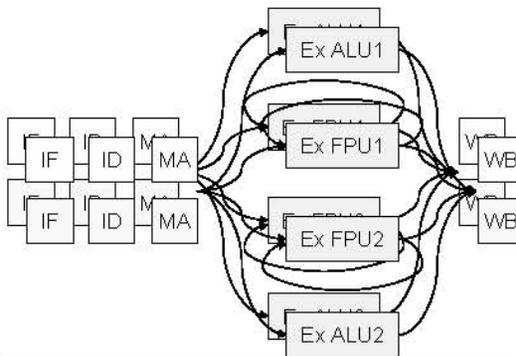


Verbesserung der Prozessor-Architektur



Weg von der klassischen SISD-Struktur ...  
 ... (Befehle werden einzeln und nur nacheinander ausgeführt.)

... hin zu Dual-Core mit Hyper-Threading



Zwei Prozessorkerne auf einem Chip - Jeder Kern verarbeitet zwei parallele Threads, damit laufen bis zu vier Threads simultan und bringen mehr, als die doppelte Performance.

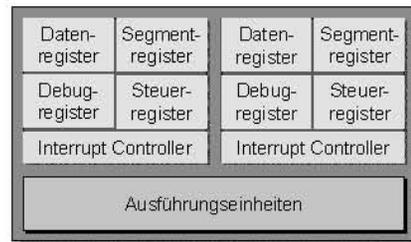
Hyper-Threading Technology (HTT)



Prozessor ohne Hyper-Threading Technology



Prozessor mit Hyper-Threading Technology

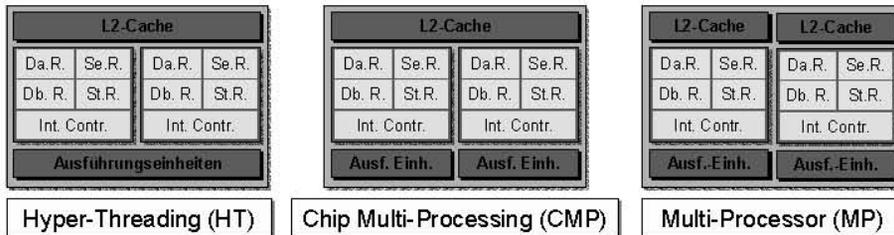


Multi-Processing - Möglichkeiten

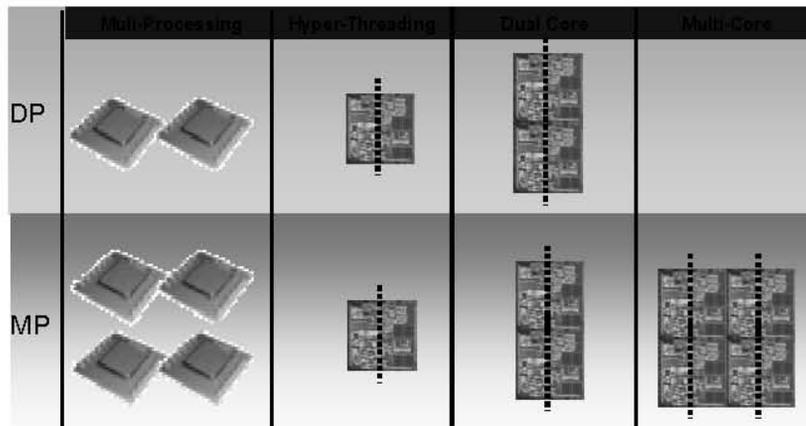
Hyper-Threading und Chip Multi-Processing sind unterschiedliche Ausführungen der Multithreading Technologie

- Hyper-Threading entspricht einem Ein Prozessor System, in dem Ressourcen geteilt werden (doppelte Anzahl der Register)
- Chip Multi-Processing entspricht einem Dual Prozessor System. Zustände (Register) und Ressourcen sind doppelt vorhanden und der L2 Cache wird geteilt

Hyper-Threading und Chip Multi-Processing können den gleichen Code nutzen



Vergrößerung der Parallelität pro Sockel



Multiprocessor vs. Multicore

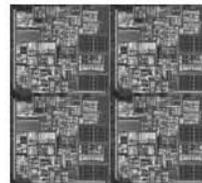
Multiprocessor

- Separate Chips, die extern mit einem Bussystem verbunden sind
- Erfordern spezielle Softwareunterstützung
- Hohe Verlustleistung



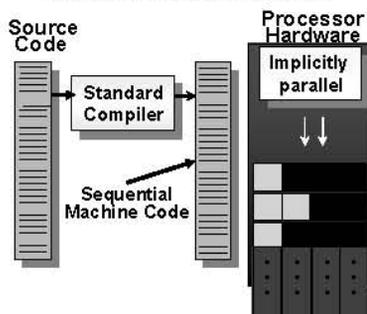
Multicore

- Die Kerne sind auf dem Chip direkt miteinander verbunden
- Threads und andere Prozesse werden automatisch parallel abgearbeitet
- Nur geringe Erhöhung der Verlustleistung

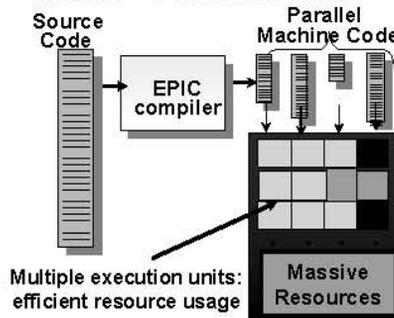


Verbesserung der Prozessor-Architektur

Traditional IA Architecture



Itanium™ EPIC Architecture



- Höhere Performance mit 64-Bit EPIC compilierten Applikationen
- Compiler basierende Optimierung und parallele Prozessor-Ressourcen bewirken eine effizientere Performance-Steigerung



Forderungen an die Prozessorentwicklung

- **Isolation der Anwenderprogramme**
  - gegeneinander und gegenüber dem Betriebssystem
  - fehlerhafte Programme dürfen nur sich selbst beeinflussen
- **Speicherverwaltung**
  - Zuweisung freier Speicherbereiche für Code, Daten und Stack
  - Verhinderung von Fragmentierung
  - Unterstützung von *virtuellem* Speicher
- **dynamische Adressbindung (*Linking*)**
  - Code zur Laufzeit an den aktuell belegten Adressbereich binden
- **wesentliche Erhöhung der Verarbeitungsleistung**
  - Verwaltungs-Overhead Betriebssystem
  - Multitasking

