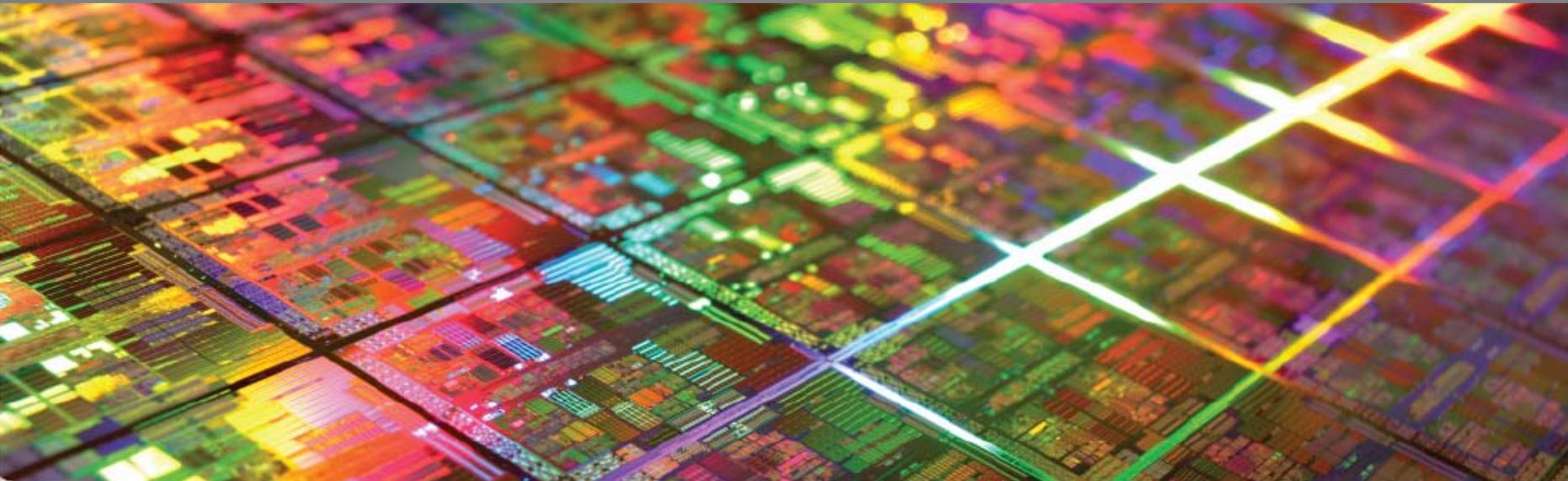


Rechnerstrukturen

Vorlesung im Sommersemester 2010

Prof. Dr. Wolfgang Karl

Fakultät für Informatik – Lehrstuhl für Rechnerarchitektur und Parallelverarbeitung



Vorlesung Rechnerstrukturen

- Kapitel 1: Grundlagen
 - 1.4 Bewertung der Leistungsfähigkeit

Bewertung der Leistungsfähigkeit

■ Ziele

- Auswahl der Rechenanlage
- Veränderung der Konfiguration einer bestehenden Anlage
- Entwurf von Anlagen

Benchmarks

■ Standardisierte Benchmarks

■ Standardisierungsorganisationen

■ SPEC (Standard Performance Evaluation Corporation)

■ Gegründet 1988, <http://www.spec.org>

- Zusammenschluss von mehr als 40 Firmen (Rechnerhersteller)
- Festlegung von Richtlinien für eine gemeinsame Rechnerbewertung

■ SPEC's Structure

- *“SPEC is a non-profit corporation whose membership is open to any company or organization that is willing to support our goals (and pay our nominal dues). Originally just a bunch people from workstation vendors devising CPU metrics, SPEC has evolved into an umbrella organization encompassing three diverse groups.”*

Benchmarks

■ SPEC (Standard Performance Evaluation Corporation)

■ Struktur:

■ Open Systems Group (OSG)

- CPU: SPECmarks CPU Benchmarks
- Java: Client- und Serverseitige Benchmarks: JVM98, JVM2008, JBB2000, JBB2005, jAppServer Java Enterprise Application Server benchmarks
- Mail: SPECmail2001, Consumer Internet Service Provider (ISP) mail server benchmark
- Power: SPECpower_ssj2008, der SPEC benchmark zur Evaluierung der Energieeffizienz für Server
- SIP: SPEC Benchmark zum Vergleich von Servern, die das Session Initiation Protokoll (SIP) verwenden
- SFS: SFS93 (LADDIS), SFS97, SFS97_R1, and SFS2008
- Virtualization: Entwicklung der ersten Generation eines SPEC Benchmarks zum Vergleich der Virtualisierungsleistung für Data Centers
- WEB: WEB96, WEB99, WEB99_SSL, and WEB2005, die web server benchmarks.

Benchmarks

- **SPEC (Standard Performance Evaluation Corporation)**
- Struktur:
- The High-Performance Group (HPG)
 - Benchmark Suite, die HPC Anwendungen repräsentieren
 - Zielarchitekturen: symmetrische Multiprozessorsysteme, Workstation Cluster, Parallelrechner mit verteiltem Speicher, Vektorrechner
- The Graphics and Workstation Performance Group (GWPG):
 - graphics and workstation performance benchmarks and reporting procedures
 - SPECapc (Application Performance Characterization): CAD/CAM, Digital Content Creation, Visualisierungsanwendungen
 - SPECgpc (Graphics Performance Characterization group): Benchmarks zur Leistungsbewertung von Graphicsystemen, die unter OpenGL oder anderen APIs laufen, SPECviewperf(r)

Benchmarks

- **SPEC (Standard Performance Evaluation Corporation)**
- SPEC CPU Benchmarks
- Strenge, genau festgelegte Regeln
- Ab CPU95: vollautomatische Messung und Protokollierung
- Regelmäßige Aktualisierungen (CPU92, CPU95, CPU2000)
 - Laufzeiten werden zu kurz
 - Caches werden größer: größere Datensätze
 - Mehr Praxisnähe: Programme mit schlechterer Datenlokalität
- SPEC CPU 2006
 - 12 nichtnumerische Programme in C/C++ (CINT2006): [SPEC CINT2006 Benchmarks.htm](#)
 - 14 numerische Programme in FORTRAN/C (CFP2006): <http://www.spec.org/cpu2006/CFP2006/>
 - Referenzmaschine:
 - Historisches Sun System, eine "Ultra Enterprise 2 mit einem 296 MHz UltraSPARC II Prozessor

Benchmarks

- **SPEC (Standard Performance Evaluation Corporation)**
- SPEC CPU 2006

	Geschwindigkeit	Durchsatz
Aggressive Optimierung	SPECint2000	SPECint_rate2000
	SPECfp2000	SPECfp_rate2000
Konservative Optimierung	SPECint_base2000	SPECint_rate_base2000
	SPECfp_base2000	SPECfp_rate_base2000

Benchmarks

- **SPEC (Standard Performance Evaluation Corporation)**
- SPEC CPU 2006 Benchmark-Metrik Geschwindigkeit

$$\text{SPECratio} = \frac{\text{Referenzzeit}_x}{\text{Laufzeit}_x \text{ auf Testsystem}} \quad \text{Benchmark } x$$

- Endwerte: je ein geometrisches Mittel der SPECratio's über alle CINT2006 und CFP2006 Benchmarks
 - SPECint2006, SPECfp2006
 - Aggressive, individuelle Optimierungen erlaubt
 - SPECint_base2006, SPECfp_base2006
 - Nur mit konservativer Standardoptimierung
 - Identische Compileroptionen für alle Programme

Benchmarks

- **SPEC (Standard Performance Evaluation Corporation)**
- SPEC CPU 2006 Benchmark-Metrik Geschwindigkeit
 - Warum Geometrisches Mittel

$$\sqrt[n]{\prod_{i=1}^n \text{Execution time ratio}_i}$$

mit *Execution time ratio*:
 Ausführungszeit der Programms i
 einer Last von n Programmen,
 normalisiert bezüglich der
 Referenzmaschine

- Eigenschaft des geometrischen Mittels:

$$\frac{\text{geometrisches Mittel } (X_i)}{\text{geometrisches Mittel } (Y_i)} = \text{geometrisches Mittel} \left(\frac{X_i}{Y_i} \right)$$

Geometrisches Mittel ist konsistent, unabhängig von Referenzmaschine!

Benchmarks

- **SPEC (Standard Performance Evaluation Corporation)**
- SPEC CPU 2006 Benchmark-Metrik Durchsatz

$$\text{SPECrate}_x = \frac{\text{Sekunden pro Stunde}}{\text{Laufzeit}_x \text{ von } n_x \text{ Kopien auf Testsystem}} \times \frac{\text{Referenzzeit}_x}{\text{längste Laufzeit}}$$

- Endwerte: je ein geometrisches Mittel der SPECrate's über alle CINT2000 bzw. CFP2000 Benchmarks
 - SPECint_rate2000, SPECfp_rate2000
 - SPECint_base2000, SPECfp_base2000
 - n_x kann frei gewählt werden, muss aber dokumentiert werden

Benchmarks

■ SPEC (Standard Performance Evaluation Corporation)

Processor	Alpha 21364	AMD Athlon XP	HP PA-8700	IBM Power 4+	Intel Intelium 2	Intel XeonMP	Intel Xeon	MIPS R14000	Sun UltraSPARC III
System or Motherboard	Alpha GS1280/7	ASUS A7N8X	HP9000 C3750	pSeries 650 6M2	HP RX2600	Dell PwrEdg 6650	Dell Prec. 350	SGI 3200	Sun Blade 2050
Clock Rate	1.15GHz	2.17GHz	870MHz	1.45GHz	1.0GHz	2.0GHz	3.06GHz	600MHz	1.05GHz
External Cache	None	None	None	16MB	None	None	None	8MB	8MB
164.gzip	583	1,026	588	673	583	758	1,138	322	433
175.vpr	822	653	688	902	704	625	606	572	460
176.gcc	859	755	906	914	1,014	1,100	1,236	445	577
181.mcf	712	420	494	1,391	834	599	773	783	659
186.crafty	982	1,292	751	884	781	712	1,179	502	558
197.parser	514	905	495	381	660	778	1,025	409	488
252.eon	958	1,483	592	1,150	1,004	920	1,387	507	527
253.perlbnk	768	1,306	619	712	815	952	1,381	367	540
254.gap	636	1,059	339	936	680	722	1,417	308	372
255.vortex	1,094	1,608	1,196	1,428	1,193	1,118	1,658	679	738
256.bzip2	824	840	534	965	759	712	856	493	629
300.twolf	1,018	887	911	1,198	880	1,009	900	645	570
SPECint_base2000	795	960	642	909	810	816	1,085	483	537
168.wupside	883	1,131	446	1,532	1,003	816	1,406	434	659
171.swim	3,590	1,006	931	1,417	3,205	848	1,837	529	980
172.mgrid	708	799	621	850	1,720	449	1,047	379	487
173.applu	1,518	654	702	979	2,033	496	1,168	381	310
177.mesa	928	1,103	694	737	642	814	1,165	425	543
178.galgel	2,105	738	1,603	3,186	2,505	1,200	1,536	1,398	1,713
179.art	2,014	495	670	1,864	4,226	1,147	716	1,436	9,389
183.equake	519	730	413	2,098	1,871	449	1,291	347	645
187.facerec	1,105	1,008	430	1,515	1,152	762	1,315	647	958
188.ammp	735	587	553	923	788	729	644	573	509
189.lucas	1,522	853	448	1,306	1,206	682	1,522	442	371
191.fma3d	1,019	850	404	898	747	551	1,089	306	400
200.sixtrack	469	538	471	621	894	376	564	298	366
301.aspi	1,242	705	696	966	678	695	833	406	471
SPECfp_base2000	1,124	776	600	1,221	1,356	677	1,092	499	701

Benchmarks

■ SPEC (Standard Performance Evaluation Corporation)

Processor	AMD 1-core Opteron 854	Intel 1-core Xeon	AMD 2-core Opteron 8224SE	Intel 2-core Xeon 5160	AMD 4-core Opteron 8360SE	Intel 4-core Xeon X7350	Intel 4-core Core 2 Quad QX9650
Bit-width	32/64-bit	32/64-bit	32/64-bit	32/64-bit	32/64-bit	32/64-bit	32/64-bit
Cores/chip x Threads/core	1 x 1	1 x 2	2 x 1	2 x 1	4 x 1	4 x 1	4 x 1
Clock Rate	2.80GHz	3.80GHz	3.20GHz	3.03GHz	2.50GHz	2.93GHz	3.00GHz
Cache: L1-L2-L3 - I/D or Unified	64K/64K - 1M - N/A	12K/16K - 2M - N/A	2 x 64K/64K - 2 x 1M - NA	2 x 32K/32K - 4M - NA	4 x 64K/64K - 4 x 512K - 2M	4 x 32K/32K - 2 x 4M - NA	4 x 32K/32K - 2 x 6M - NA
Execution Rate/Core	3 instructions	3 instructions	3 instructions	1 complex + 3 simple	3 instructions	1 complex + 3 simple	1 complex + 3 simple
Pipeline Stages	12 int / 17 fp	31	12 int / 17 fp	14	12 int / 17 fp	14	14
Out of Order	72	126	72	96	72	96	96
Memory bus	6.4 GB/s	800 MHz	10.6 GB/s	1333 MHz	10.6 GB/s	1066 MT/s	1333 MHz
Package	uPGA 940	LGA-775	LGA-1207	LGA-771	LGA-1207	LGA-771	LGA-775
IC Process	90nm 9M	90nm 7M	90nm 9M	65nm 8M	65nm 8M	65nm 8M	45nm
Die Size	106mm ²	109mm ²	227mm ²	143mm ²	283mm ²	2 x 143mm ²	2 x 107mm ²
Transistors	120M	169M	233M	291M	463M	2 x 291M	2 x 410M
List Price (intro)	\$1,514	\$903	\$2,149	\$851	N/A	\$2,301	\$999
Power (Max)	93W	110W	120W	80W	120W	130W	130W
Availability	3Q05	3Q05	3Q07	3Q06	1Q08	3Q07	4Q07
Scalability	2-4 chips	1-2 chips	1-4 chips	1-2 chips	2-4 chips	1-4 chips	1 chip
SPECint/fp2006 [cores]	11.2/12.1 [2]	11.4/11.7 [2]	14.1/14.2 [8]	19.7*/18.3* [4]	N/A	21.7*/18.9* [16]	22.3*/21.4* [4]
SPECint/fp2006_rate [cores]	41.4/45.6 [4]	20.9/18.8 [2]	105/96.7 [8]	60.8/45.1 [4]	163/149 [16]	184*/108* [16]	69.0*/49.9 [4]

Quelle: MDR, Microprocessor Report, January 2008

Benchmarks

■ SPEC (Standard Performance Evaluation Corporation)

Processor	Intel Itanium 2 9050	Intel Itanium 9150M	IBM Power6	IBM Power5+	Fujitsu SPARC64 VI	Sun UltraSPARC IV+	Sun UltraSPARC T2
Bit-width	64-bit	64-bit	64-bit	64-bit	64-bit	64-bit	64-bit
Cores/chip x Threads/core	2 x 2	2 x 2	2 x 2	2 x 2	2 x 2	2 x 1	8 x 8
Clock Rate	1.60GHz	1.67GHz	4.70GHz	2.20GHz	2.40GHz	1.95GHz	1.40GHz
Cache: L1-L2-L3 I/D or Unified	2 x 16K/16K - 1M/256K - 12M(on)	2 x 16K/16K - 1M/256K - 12M(on)	2 x 64K/64K - 2 x 4M - 32M(off)	2 x 64K/32K - 1.92M - 36M(off)	2 x 128K/128K - 5M - NA	2 x 64K/64K - 2M - 32M(off)	8 x 8K/16K - 4M - NA
Execution Rate/Core	6 issue	6 issue	7 issue	5 issue	4 issue	4 issue	16 issue
Pipeline Stages	8 stages	8 stages	13 stages	15 stages	15 stages	14 stages	8 int / 12 fp
Out of Order	None	None	"Limited"	200	64	None	None
Memory B/W	8.5GB/s	10.6GB/s	75GB/s	12.8GB/s	8GB/s	4.8GB/s	42.7GB/s
Package	mPGA-700	mPGA-700	N/A	MCM-5370 pins	412 I/O pins	FC-LGA 1368	1831 pins
IC Process	90nm 7M	90nm 7M	65nm 10m	90nm 10m	90nm 10M	90nm 9M	65nm
Die Size	596mm ²	596mm ²	341mm ²	245mm ²	421mm ²	335mm ²	342mm ²
Transistors	1.72 billion	1.72 billion	790 million	276 million	540 million	295 million	503 million
List Price (intro)	\$3,692	\$3,692	N/A	N/A	N/A	N/A	N/A
Power (Max)	104W	104W	~100W	100W	120W	90W	84W
Availability	3Q06	4Q07	2Q07	4Q05	2Q07	3Q06	3Q07
Scalability	1-64 chips	8-128 chips	2-32 chips	1-32 chips	4-64 chips	1-72 chips	1 chip
SPECint/fp2006 [cores]	14.5/17.3 [2]	N/A	17.8/18.7 [1]	10.5/12.9 [1]	9.7/11.1 [32]	N/A	N/A
SPECint/fp2006_rate [cores]	1534/1671 [128]	1832/N/A [128]	420/379 [16]	197/229 [16]	1111/1160 [128]	1120/N/A [144]	73.1/58.1

Quelle: MDR, Microprocessor Report, January 2008

Benchmarks

■ SPEC (Standard Performance Evaluation Corporation)

Processor	Intel 2-core Xeon X5270 ¹	AMD 2-core Opteron 8224SE	Intel 4-core Xeon W5590	AMD 4-core Opteron 8393SE ²	Intel 4-core Xeon X5570 ³	AMD 6-core Opteron 8439SE	Intel 6-core Xeon X7460 ⁴
Bit-Width	32/64-bit	32/64-bit	32/64-bit	32/64-bit	32/64-bit	32/64-bit	32/64-bit
Cores/Chip x Threads/core	2 x 1	2 x 1	4 x 2	4 x 1	4 x 2	6 x 1	6 x 1
Clock Rate	3.50GHz	3.20GHz	3.33GHz	3.10GHz	2.93GHz	2.80GHz	2.67GHz
Cache: L1-L2-L3 - I/D or Unified	2 x 32K/32K - 6M - NA	2 x 64K/64K - 2 x 1M - NA	4 x 32K/32K - 4 x 256K - 8M	4 x 64K/64K - 4 x 512K - 6M	4 x 32K/32K - 4 x 256K - 8M	6 x 64K/64K - 6 x 512K - 6M	6 x 32K/32K - 3 x 3M - 16M
Execution Rate/Core	1 cmplx + 3 simple	3 instructions	1 cmplx + 3 simple	3 instructions	1 cmplx + 3 simple	3 instructions	1 cmplx + 3 simple
Pipeline Stages	14	12int / 17fp	16	12int / 17fp	16	12int / 17fp	14
Out of Order	96	72	128	72	128	72	96
Memory Bus	1333MHz	10.6GB/s	16GB/s	8GB/s	16GB/s	17GB/s	1064MHz
Package	LGA-771	LGA-1207	LGA-1366	LGA-1207	LGA-1366	LGA-1207	LGA-771
IC Process / Metal Layers	45nm / 10M	90nm / 10M	45nm / 10M	45nm / 10M	45nm / 10M	45nm / 10M	45nm / 10M
Die Size	107mm ²	227mm ²	263mm ²	258mm ²	263mm ²	346mm ²	503mm ²
Transistors	410M	233M	731M	758M	731M	904M	1900M
List Price (Intro)	\$1,172	\$2,149	\$1,600	\$2,649	\$1,386	\$2,649	\$2,729
Power (Max)	80W	120W	130W	137W (TDP)	95W	137W (TDP)	130W
Availability	3Q08	3Q07	3Q09	2Q09	1Q09	3Q09	4Q08
Scalability	1-2 chips	1-4 chips	1-2 chips	1-8 chips	1-2 chips ³	1-8 chips	1-4 chips
SPECint/fp2006 [cores]	26.5/25.5 [4]	14.1/14.2 [8]	34.2 ^a /40.4 ^a [8]	19.7/23.6 [8]	32.1 ^a /45.0 [8]	18.3/23.3 [12]	22.0/22.3 [24]
SPECint/fp2006_rate [cores]	85.3/57.7 [4]	105/96.7 [8]	255/204 [8]	232/204 ² [16]	240/197 [8]	629/473 [48]	274 ⁴ /142 [24]
x86 Codename	Wolfdale	Santa Rosa	Gainestown	Shanghai	Gainestown	Istanbul	Dunnington
Microarchitecture	Core	K8	Nehalem	K10	Nehalem	K10	Core

Quelle: MDR, Microprocessor Report, January 2010

Benchmarks

■ SPEC (Standard Performance Evaluation Corporation)

Processor	Intel Itanium 2 9050	Intel Itanium 9150M ⁵	IBM POWER5+	IBM POWER6+ ⁶	Fujitsu SPARC64 VI	Fujitsu SPARC64 VII	Sun UltraSPARC T2+ ⁷
Bit-Width	64-bit	64-bit	64-bit	64-bit	64-bit	64-bit	64-bit
Cores/Chip x Threads/core	2 x 2	2 x 2	2 x 2	2 x 2	2 x 2	4 x 2	8 x 8
Clock Rate	1.60GHz	1.67GHz	2.20GHz	5.00GHz	2.40GHz	2.52GHz	1.60GHz
Cache: L1-L2-L3 - I/D or Unified	2 x 16K/16K - 1M/256K - 12M(on)	2 x 16K/16K - 1M/256K - 12M(on)	2 x 64K/32K - 1.92M - 36M(off)	2 x 64K/64K - 2 x 4M - 32M(off)	2 x 128K/128K - 6M - N/A	4 x 64K/64K - 6M - N/A	8 x 16K/8K - 4M - NA
Execution Rate/Core	6 Issue	6 Issue	5 Issue	7 Issue	4 Issue	4 Issue	16 Issue
Pipeline Stages	8	8	15	13	15	15	8int / 12fp
Out of Order	None	None	200	"Unlimited"	64	64	None
Memory Bus	8.5GB/s	10.6GB/s	12.8GB/s	75GB/s	8GB/s	8GB/s	42.7GB/s
Package	mPGA-700	mPGA-700	MCM-5370 pins	N/A	412 I/O pins	412 I/O pins	1831 pins
IC Process / Metal Layers	90nm / 7M	90nm / 7M	90nm / 10M	65nm / 10M	90nm / 10M	65nm / 11M	65nm / 8M
Die Size	596mm ²	596mm ²	245mm ²	341mm ²	421mm ²	400mm ²	342mm ²
Transistors	1.72B	1.72B	276M	790M	540M	600M	503M
List Price (Intro)	\$3,692	\$3,692	N/A	N/A	N/A	N/A	N/A
Power (Max)	104W	104W	100W	>100W	120W	135W	95W ⁷
Availability	3Q06	4Q07	4Q05	4Q08 ⁶	2Q07	3Q08	3Q09
Scalability	1-64 chips	8-128 chips	1-32 chips	2-32 chips	4-64 chips	4-64 chips	1-4 chips
SPECint/fp2006 [cores]	14.5/17.3 [2]	N/A	10.5/12.9 [1]	15.8/20.1 [1]	9.7/21.7 [32]	11.5/13.0 [1]	N/A
SPECint/fp2006_rate [cores]	1534/1671 [128]	2893/N/A [256]	197/229 [16]	1866/1822 [64]	1111/1160 [128]	2088/1861 [256]	338/254 [32]
Development Status	Inactive	active	Inactive	active	Inactive	active	active

All SPEC scores are base. ^aScore measured in single-thread mode.

NOTES:

¹Higher-numbered X5272 at lower 3.40GHz frequency has a faster 1600MHz front-side bus and posts a slightly higher score (26.6) on the CINT2006 benchmark. Among Intel's dual-core processors, the 2.93GHz E7220 scales to 4 chips and consequently boasts the highest SPECrate INT/FP numbers (139/89.4).

²Posted 8-chip [32-core] SPEC intrate score of 338, but no official 8-chip SPEC fprate score.

³Although 2 chips is the design limit for an MP board, the X5570 is used in "virtual SMP" systems based on scalable blade architectures that, in principle, can go as high as 32K chips. Highest posted SPEC intrate/fprate score is for 32 chips [128 cores]: 3147/2553.

⁴Unisys ES7000 Model 7600R has a posted 96-core [16 chip] CINT2006rate number of 999 (but no other SPEC measures).

⁵Dual-core Itanium 2 9040 at 1.60GHz in SGI Altrix 4700 Bandwidth System has posted 512 chip [1024 core] CFP2006 INRate/FPRate scores of 9031/10583.

⁶POWER6+ at 5.0GHz officially introduced in April 2009, but began shipping in 4Q08.

⁷Wattage at 1.40GHz, no published update with 3Q09 upgrade to 1.6GHz frequency.

Quelle: MDR, Microprocessor Report, January 2010

Literatur

- Hennessy/Patterson: A Quantative Approach: Kap. 1.5 – 1.9

Bewertung der Leistungsfähigkeit

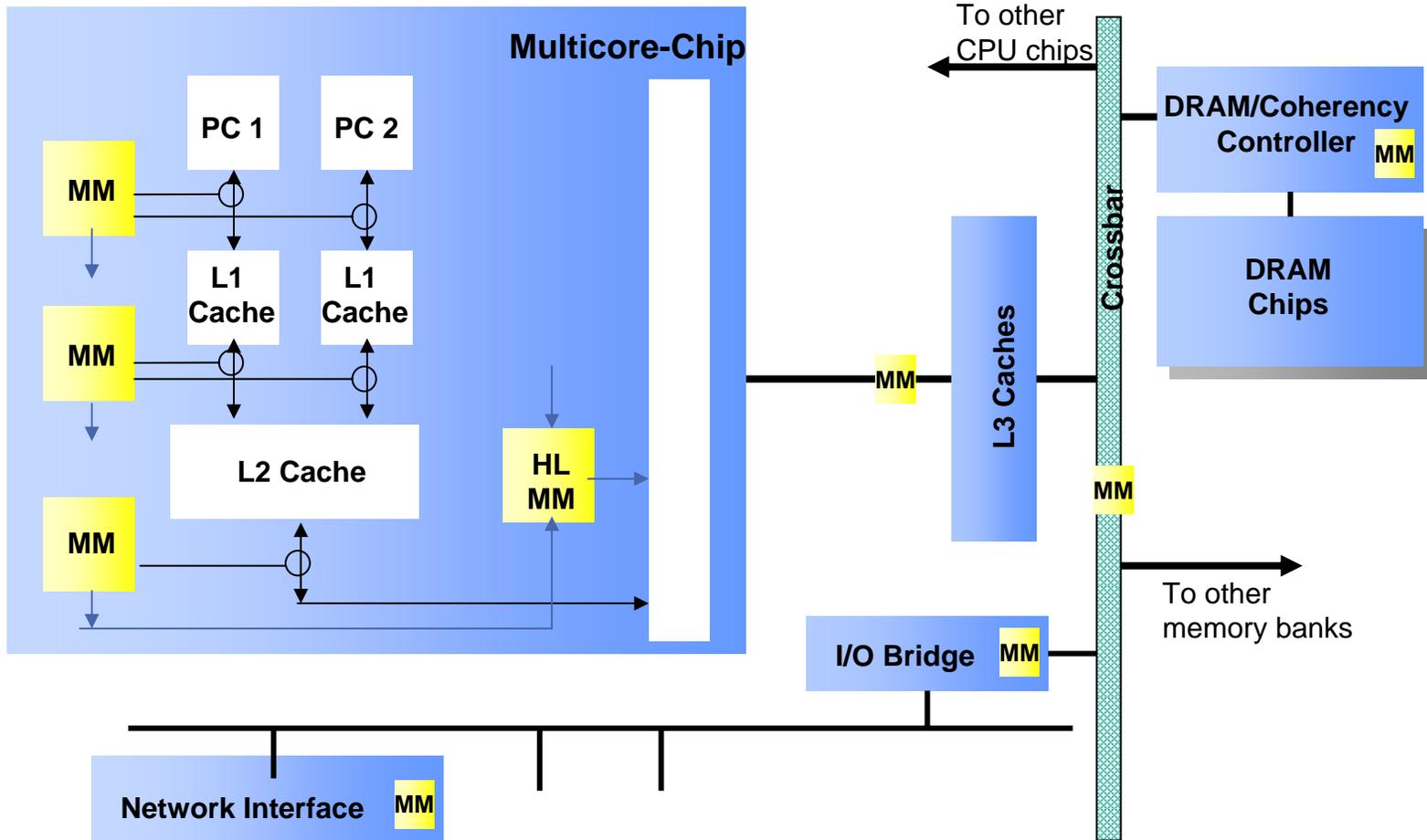
- **Messung während des Betriebs von Anlagen**
- **Monitore**
 - Aufzeichnungselemente, die zum Zweck der Rechnerbewertung die Verkehrsverhältnisse während des normalen Betriebs beobachten und untersuchen.
 - **Hardware-Monitore**
 - Unabhängige physikalische Geräte
 - Keine Beeinflussung
 - **Software-Monitore**
 - Einbau in das Betriebssystem
 - Beeinträchtigung der normalen Betriebsverhältnisse

Bewertung der Leistungsfähigkeit

- **Messung während des Betriebs von Anlagen**
- Monitore
- Aufzeichnungstechniken:
 - Kontinuierlich oder sporadisch
 - Gesamtdatenaufzeichnung (Tracing)
 - Realzeitauswertung
 - Unabhängiger Auswertungslauf (Post Processing)

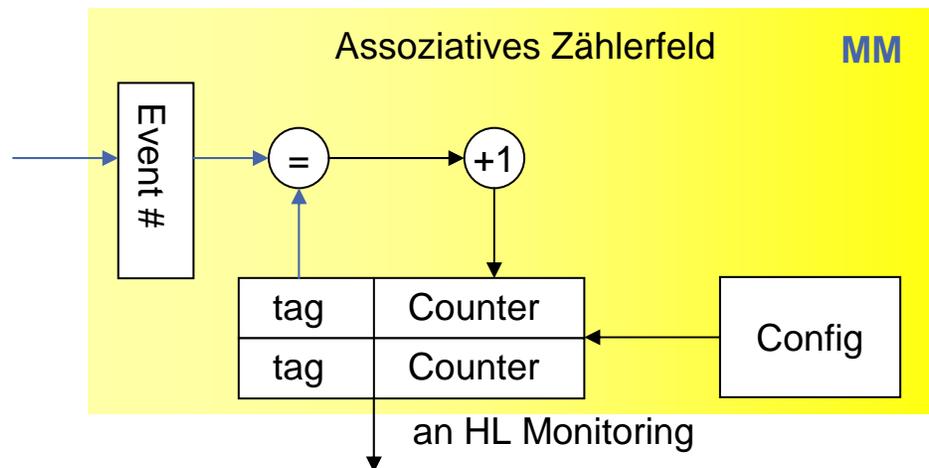
Beobachtung des Laufzeitverhaltens

- Fallstudie: CAPP Systemweite Monitoring Infrastruktur
- Knoten:



Beobachtung des Laufzeitverhaltens

- **Fallstudie: CAPP Systemweite Monitoring Infrastruktur**
- HW-Ebene: Monitorkomponenten
 - Grundlegender Aufbau eines Monitor Moduls (MM)
 - Assoziatives Zählerfeld
 - Zählen von Ereignissen
 - Korrelation von Ereignissen
 - Auslagern eines Tag/Zählerfeldes an HL Monitoring bei Zählerüberlauf / Verdrängung



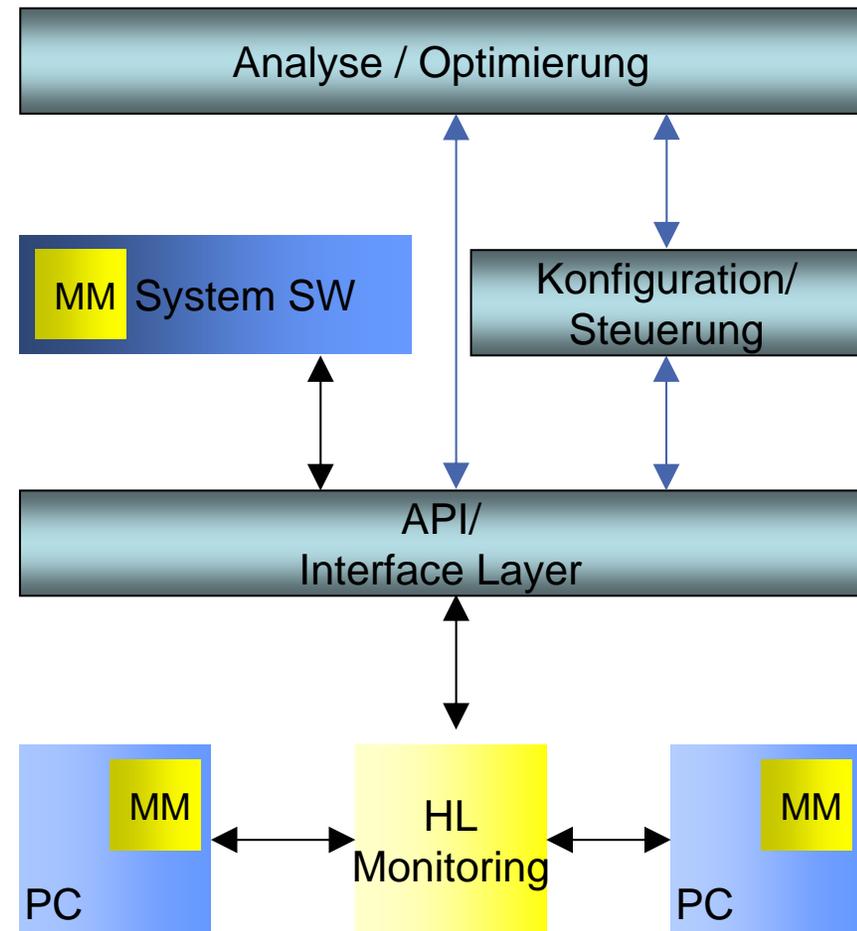
Beobachtung des Laufzeitverhaltens

■ Fallstudie: CAPP Systemweite Monitoring Infrastruktur

- Monitoring-Infrastruktur
- Monitor Module (HW, SW)

- API / Interface Layer
 - Einheitliche Schnittstelle
 - Einfache Kontrolle und Steuerung
 - Verwaltung

- Kommunikationsprotokoll
 - Leichtgewichtig und erweiterbar
 - Nachrichtenorientiert
 - Ereignisse
 - Konfiguration
 - Anstoßen der Aufzeichnung



Beobachtung des Laufzeitverhaltens

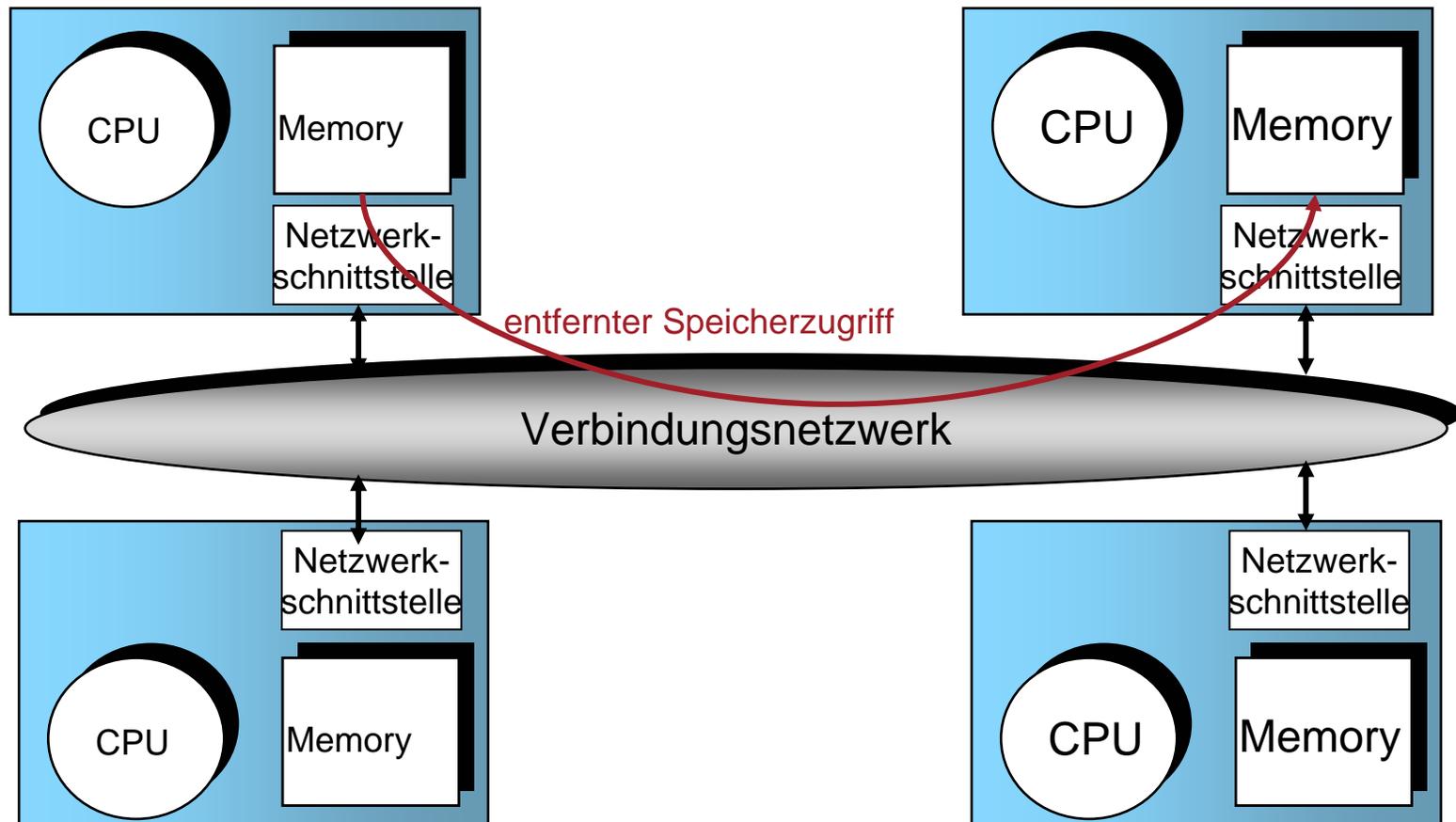
- **Fallstudie: CAPP Systemweite Monitoring Infrastruktur**
- Koordiniertes, kooperatives und systemweites Monitoring
- Merkmale
 - Geringer HW-Aufwand
 - Geringer Kommunikationsbedarf
 - Zeitliche, inhaltliche Auflösung
 - On-line Auswertung
- Auswerten der gesammelten Informationen
 - Zusammenspiel mit Werkzeugen zur Optimierung, Lastverteilung, Energieverbrauch, ...
- Herausforderungen
 - Korrelation von Ereignissen
 - Proaktivität vs. Reaktivität
 - Steuerung des adaptiven Verhaltens

Beobachtung des Laufzeitverhaltens

- **Fallstudie: CAPP Systemweite Monitoring Infrastruktur**
- Realisierungen in HW
 - SMiLE-Projekt
 - Ziel: Datenlokalitätsoptimierung – Aufspüren entfernter Speicherzugriffe in einem Cluster-System mit NUMA-Eigenschaften
 - OpenSPARC
 - HyperTransport
 - Integration in HTX-Board (FPGA) zur Beobachtung des Kommunikationsverhaltens
 - DodOrg (Digital On-Demand Computing Organism)
 - Ziel: Systemarchitektur mit Merkmalen der Selbstorganisation
 - Hierarchisches Monitor-Konzept

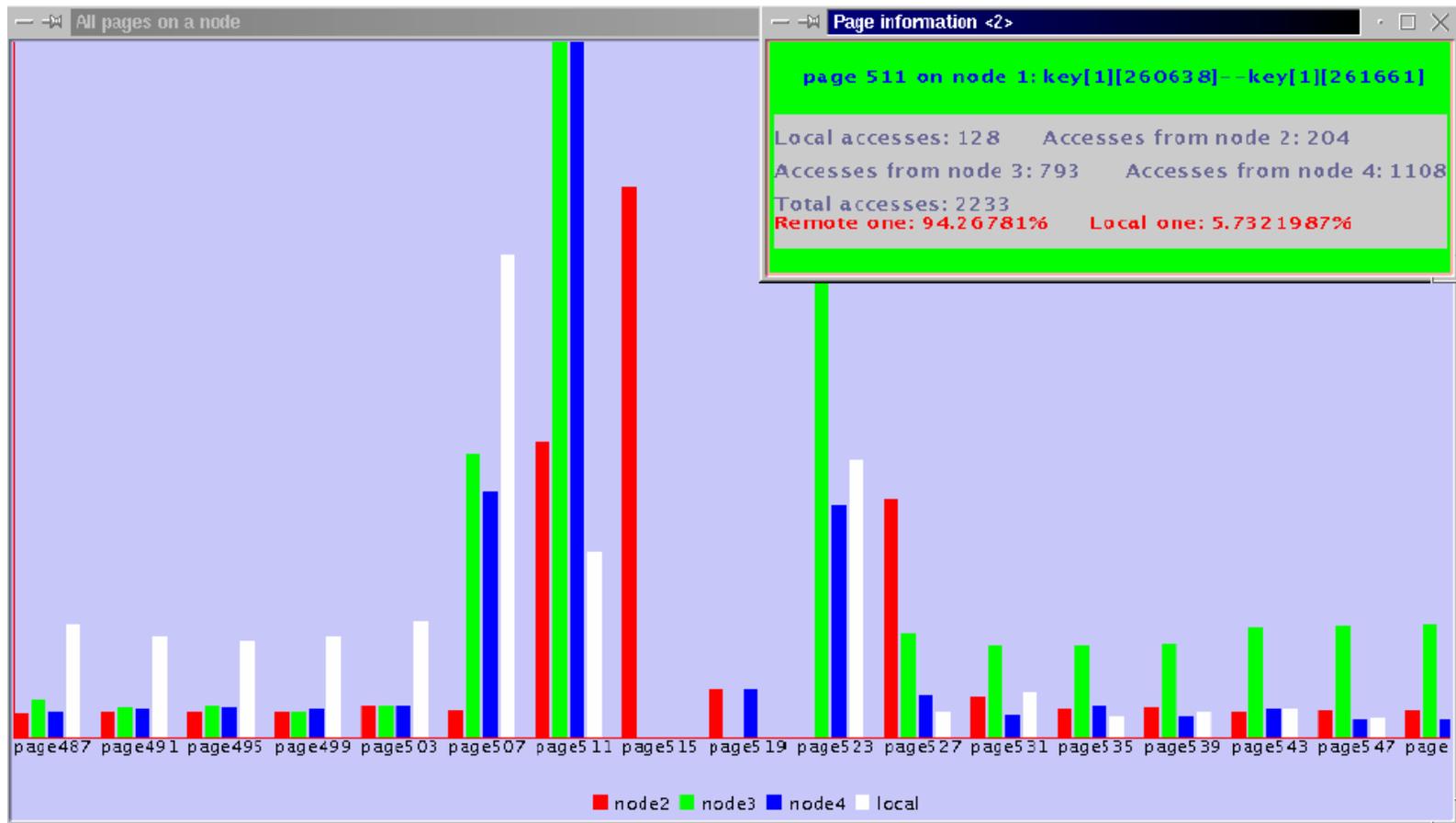
Beobachtung des Laufzeitverhaltens

- Fallstudie: CAPP Systemweite Monitoring Infrastruktur
- NUMA-System



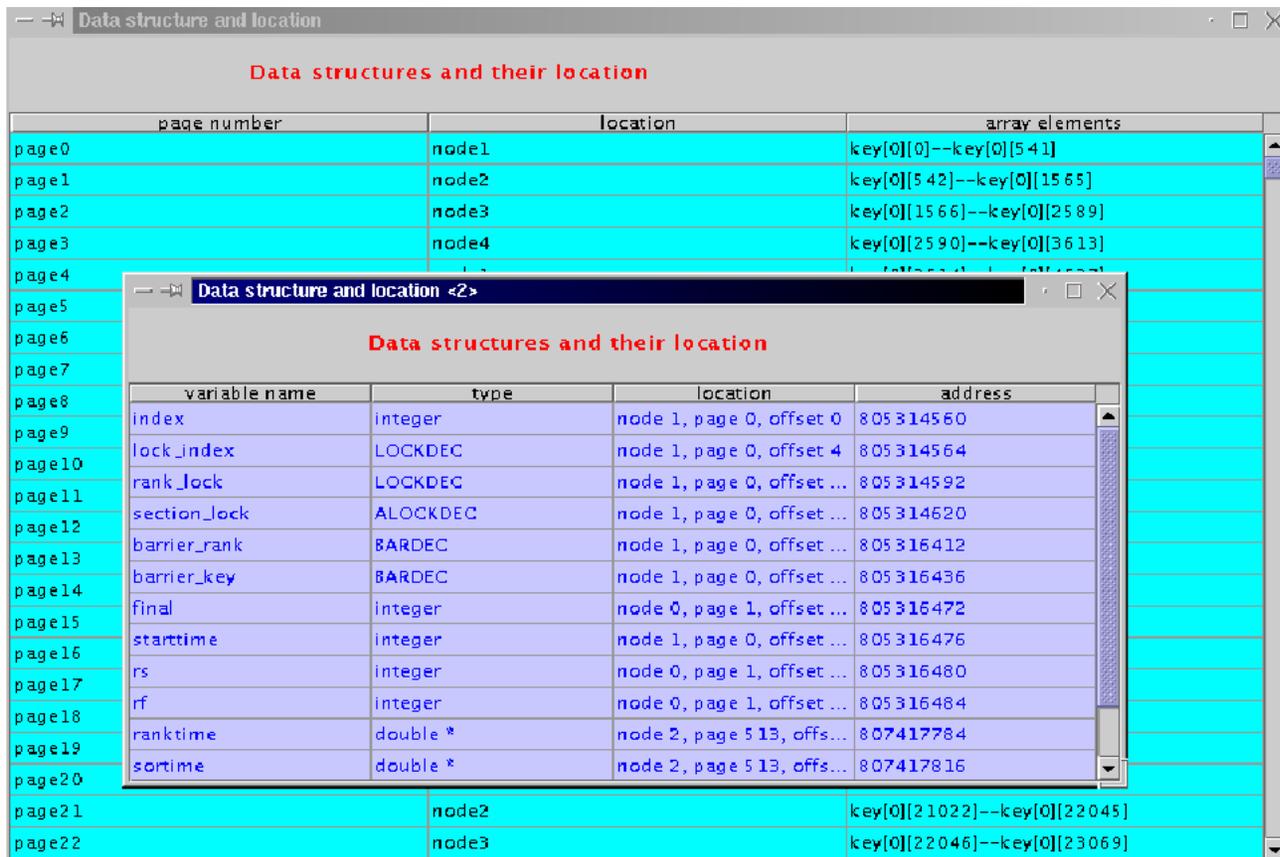
Beobachtung des Laufzeitverhaltens

- Fallstudie: CAPP Systemweite Monitoring Infrastruktur
- Visualisierung: Lokale und entfernte Speicherzugriffe



Beobachtung des Laufzeitverhaltens

- Fallstudie: CAPP Systemweite Monitoring Infrastruktur
- Visualisierung: Rückführung auf Quellcode



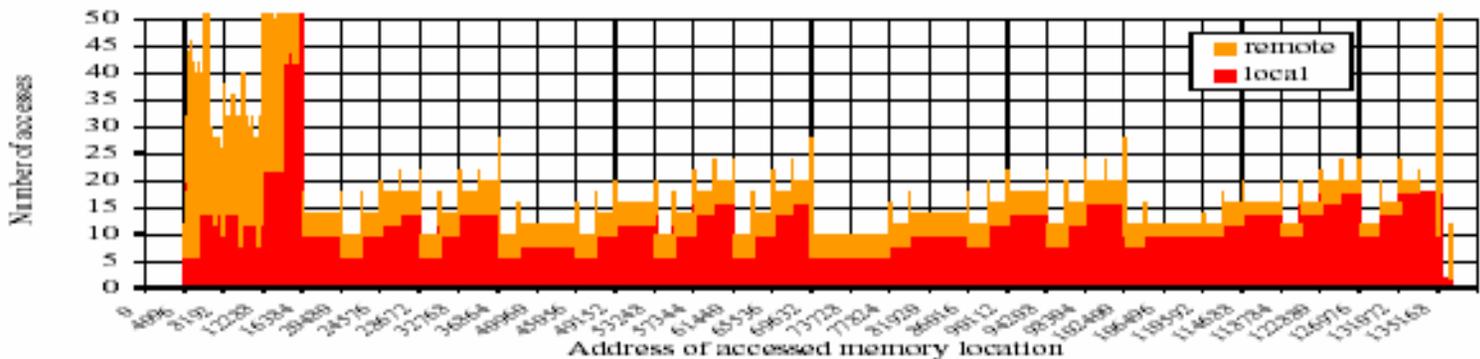
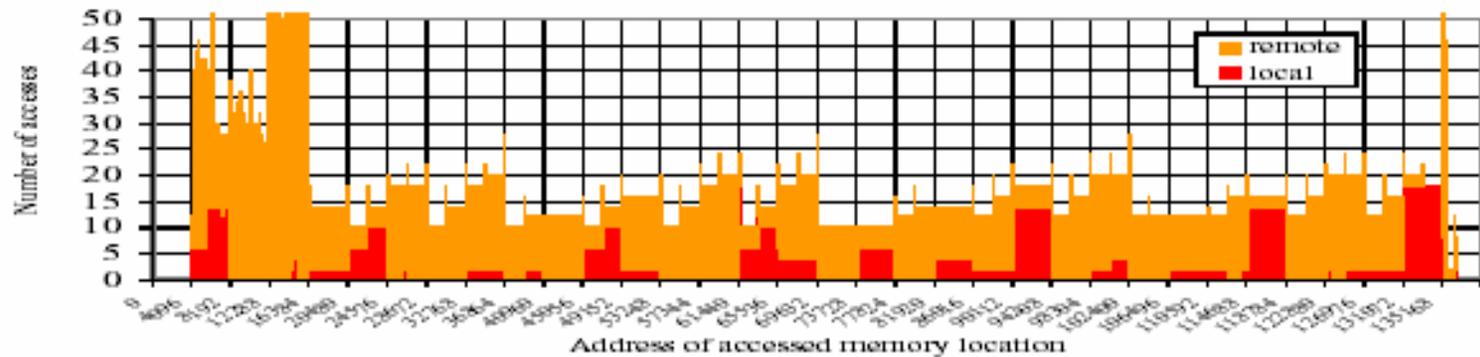
The screenshot shows a debugger window titled "Data structure and location" with a table of memory pages. An inset window shows a detailed view of variables.

page number	location	array elements
page0	node1	key[0][0]--key[0][541]
page1	node2	key[0][542]--key[0][1565]
page2	node3	key[0][1566]--key[0][2589]
page3	node4	key[0][2590]--key[0][3613]
page4		
page5		
page6		
page7		
page8		
page9		
page10		
page11		
page12		
page13		
page14		
page15		
page16		
page17		
page18		
page19		
page20		
page21	node2	key[0][21022]--key[0][22045]
page22	node3	key[0][22046]--key[0][23069]

variable name	type	location	address
index	integer	node 1, page 0, offset 0	805314560
lock_index	LOCKDEC	node 1, page 0, offset 4	805314564
rank_lock	LOCKDEC	node 1, page 0, offset ...	805314592
section_lock	ALOCKDEC	node 1, page 0, offset ...	805314620
barrier_rank	BARDEC	node 1, page 0, offset ...	805316412
barrier_key	BARDEC	node 1, page 0, offset ...	805316436
final	integer	node 0, page 1, offset ...	805316472
starttime	integer	node 1, page 0, offset ...	805316476
rs	integer	node 0, page 1, offset ...	805316480
rf	integer	node 0, page 1, offset ...	805316484
ranktime	double *	node 2, page 5 13, offs...	807417784
sortime	double *	node 2, page 5 13, offs...	807417816

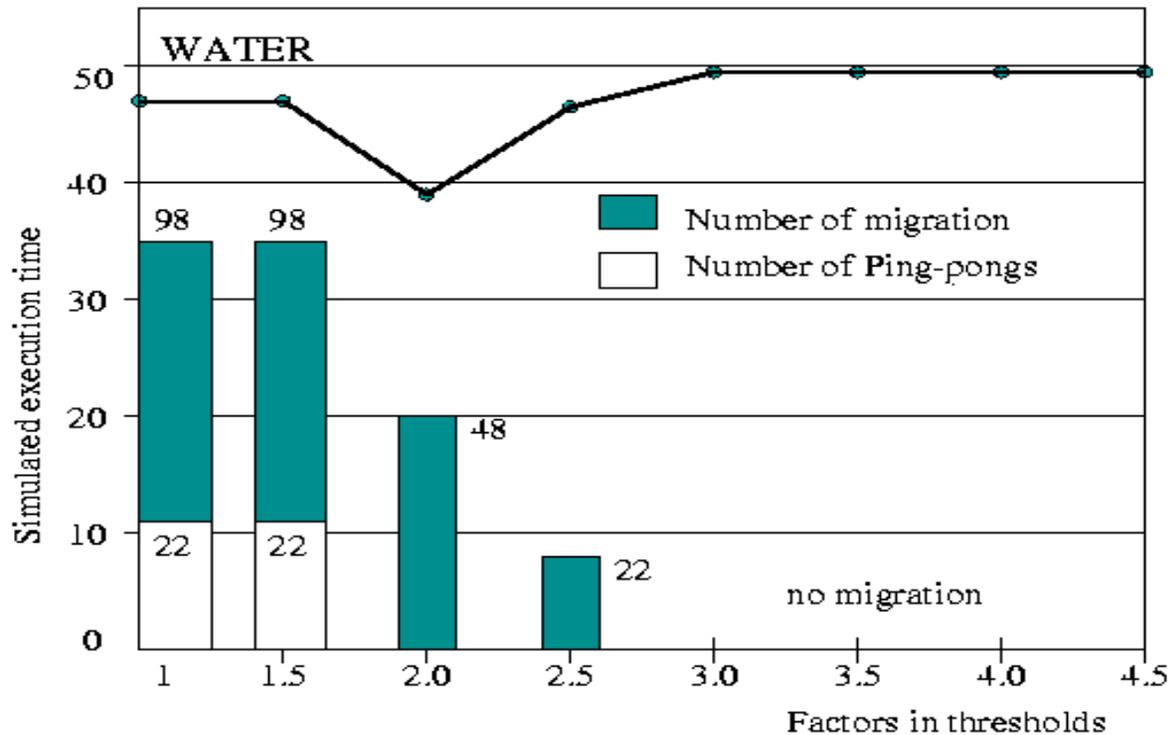
Beobachtung des Laufzeitverhaltens

- Fallstudie: CAPP Systemweite Monitoring Infrastruktur
- Beispiel LU Zerlegung: Verbesserung des Datenlokalität



Beobachtung des Laufzeitverhaltens

- Fallstudie: CAPP Systemweite Monitoring Infrastruktur
- Adaptive Optimierungskomponente:
 - Automatische Seitenmigration



Ankündigung – Informatik-Kolloquium

- **How Can Tools Keep up with the Growing Size of HPC-Systems**
- Dr. Martin Schulz, CASC, Lawrence Livermore National Laboratory, Livermore, USA
- Termin: Montag, 10.5.2010, 17:30 Uhr, Raum -101, 50.34
- *Several high-end HPC cluster systems have already surpassed the 100,000 core mark and machines with over 1,000,000 cores are on the horizon. While this promises unprecedented compute power and opens the door for new scientific discovery through advanced simulation, it comes at the price of increased complexity in both hardware and software. To deal with this complexity, users will require programming environments that scale with the machine. This includes performance analysis and debugging tools, which need to be capable of collecting, analyzing, and presenting data from all cores in a system. To satisfy these requirements we can not simply scale existing tool solutions that work on few hundred nodes; instead we require a set of new techniques that are explicitly designed and optimized for scale.*

In this talk I will show how we address this challenge in the ADEPT project at LLNL. In particular, I will focus on three of our tool sets that demonstrate different aspects necessary to achieve scalable tools: the use of hierarchical communication and online analysis in the Stack Trace Analysis Tool (STAT), the need for application specific rapid tool prototyping, as supported by the PⁿMPI infrastructure; and the design of a component infrastructure based on the Open|SpeedShop tool set

Bewertung der Leistungsfähigkeit

■ Modelltheoretische Verfahren

- Unabhängig von der Existenz eines Rechners

■ Modellbildung

- aufgrund von Annahmen über die Struktur und Betrieb eines Rechners und über die Prozesse
- Darstellung der für die Analyse relevanten Merkmale des Systems:
 - Systemkomponenten
 - Datenverkehr zwischen den Systemkomponenten
- Abstrahierung komplexer Systeme
 - Nur die interessierenden Größen werden erfasst
- Ziel:
 - Aufdecken von Beziehungen zwischen Systemparametern
 - Ermitteln von Leistungsgrößen (Auslastung von Prozessoren und Kanälen, mittlere Antwortzeiten, Warteschlangenlängen, ...)

Bewertung der Leistungsfähigkeit

- **Modelltheoretische Verfahren**
- Analytische Methoden
 - versuchen auf mathematischem Weg, Beziehungen zwischen relevanten Leistungskenngrößen und fundamentalen Systemparametern herzuleiten
 - oft nur minimaler Aufwand, aber dafür weniger aussagekräftig
- Warteschlangenmodelle
 - Leistungsanalyse von Rechensystemen
- Petrinetze
 - theoretische Untersuchungen
- Diagnosegraphen
 - Zuverlässigkeitsanalysen
- Netzwerkflussmodelle
 - Kapazitätsüberlegungen

Bewertung der Leistungsfähigkeit

■ Modelltheoretische Verfahren

■ Analytische Methoden

■ Beispiel Warteschlangenmodelle

■ Deterministische Warteschlangenmodelle

- Beispiele für Systemparameter: Rechenzeit, Gerätebedienzeit, Ankunftszeit eines Jobs
- Feste Werte → deterministische Ergebnisse für die Leistungsgrößen

■ Stochastische Warteschlangenmodelle

- Systemparameter statistisch verteilt, mit vorgegebenen Mittelwerten, Verteilungsfunktionen → statistisch verteilte Leistungsgrößen

■ Operationelle Warteschlangenmodelle

- Systemparameter: Gemessene Werte aus der Beobachtung eines Systems in einem festen Zeitintervall
- Vereinfachung der Gleichungen für die Bestimmung von Leistungsgrößen
- Relativ gute Aussagen über das Leistungsverhalten des Systems bei geeignetem gewähltem Zeitintervall

Bewertung der Leistungsfähigkeit

■ Modelltheoretische Verfahren

■ Simulation

- Vorgänge in einem Rechensystem werden nachgebildet
- Verwendung üblicher Programmiersprachen oder spezieller Simulationssprachen
- Verhalten des Simulationsmodells in Bezug auf die relevanten Parameter entspricht weitgehend dem Verhalten des realen Systems
- Ermittlung der für die Leistungsbewertung interessierenden Größen

Bewertung der Leistungsfähigkeit

- **Modelltheoretische Verfahren**
- Simulation
 - Deterministische Simulation
 - Alle an dem Modell beteiligten Größen sind exakt definiert oder berechenbar
 - Stochastische Simulation
 - Verwendung von zufallsabhängigen Größen
 - Oft: Einsatz von Zufallsgeneratoren
 - Aufzeichnungsgesteuerte Simulation
 - Verwendung von gemessenen Werten

Bewertung der Leistungsfähigkeit

■ Modelltheoretische Verfahren

■ Simulation

■ Nachteile:

- Vorbereitung und Ausführung der Simulationsmodelle zeitaufwendig und teuer
- Planung der Experimente muss sorgfältig durchgeführt werden
- Auswertung und Interpretation der Ergebnisse nicht immer einfach

■ Beispiel:

- SimpleScalar Tool Set (<http://www.simplescalar.com>)
- „Standard“-Werkzeug zur Simulation von superskalaren Mikroprozessoren

Bewertung der Leistungsfähigkeit

- **Modelltheoretische Verfahren**
- **Modellbildung**
 - Simulation vs. Analytische Methoden
 - Simulation:
 - Realistischere Annahmen über das System bei der Simulation
 - Berücksichtigung vieler verschiedener Systemgrößen
 - Abdeckungen verschiedener Anwendungsbereiche
 - Können sich gegenseitig gut ergänzen

Bewertung der Leistungsfähigkeit

■ Zusammenfassung

	Auswertung von Hardwaremaßen und Parametern	Laufzeitmessungen bestehender Programme	Messungen während des Betriebs der Anlagen	Modell-theoretische Verfahren
Rechnerauswahl	Maßzahlen für die Operationsgeschwindigkeit Kernprogramme	Benchmarks		
Rechner-„Tuning“			Hardware-Monitore Software-Monitore	
Rechnerentwurf				Analytische Methoden Simulation

Vorlesung Rechnerstrukturen

- Kapitel 1: Grundlagen
 - 1.5 Parallelverarbeitung

Parallelverarbeitung

- Sequentielle Sichtweise
 - Berechnungen werden schrittweise oder seriell ausgeführt
 - Algorithmen sind als Folge von Berechnungsschritten organisiert
 - Sequentielle Programme
 - Schrittweise Ausführung einer Folge von Befehlen auf einer sequentiellen Maschine
- Beschleunigung der Ausführung
 - Erhöhung der Taktfrequenz
 - Parallele Ausführung der Aufgaben

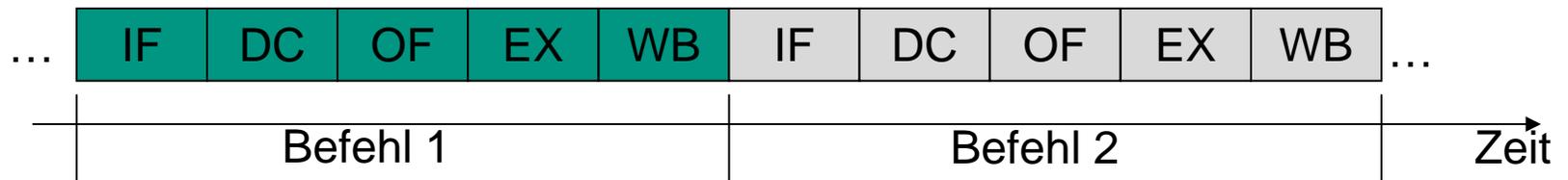
Parallelverarbeitung

- Algorithmenentwurf
 - Muss viele unabhängige Operationen umfassen
- Struktur der Programmiersprache
 - Muss die Spezifikation oder die automatische Identifizierung paralleler Operationen ermöglichen
- Rechnerarchitektur
 - Gleichzeitige Ausführung paralleler Operationen

Parallelverarbeitung

■ Sequentieller Rechner

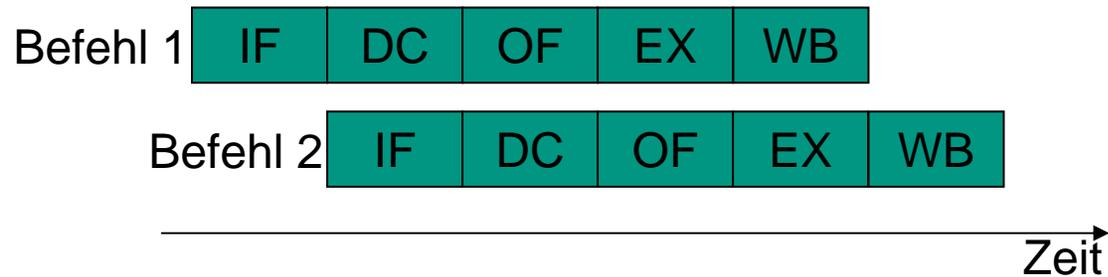
- Sequentielle Ausführung einer Folge von Maschinenbefehlen
- Maschinenbefehlszyklus



IF: Befehl holen
 DK: Befehl dekodieren
 OF: Operanden bereitstellen
 EX: Befehl ausführen
 WB: Ergebnis speichern

Parallelverarbeitung

- Pipelining des Maschinenbefehlszyklus
 - Überlappte Ausführung der Phasen des Maschinenbefehlszyklus

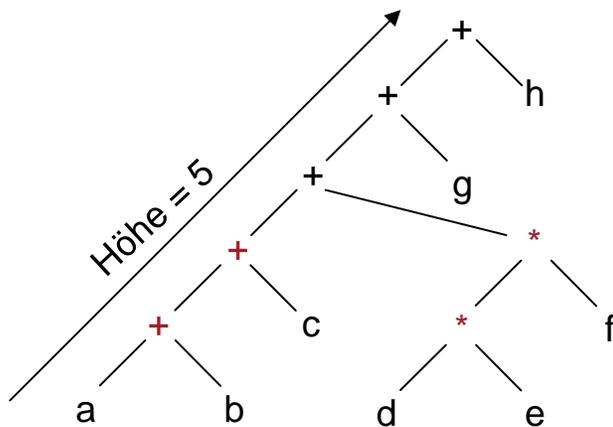


Parallelverarbeitung

- Gleichzeitige Ausführung von Operationen
- Beispiel: Auswertung des arithmetischen Ausdrucks

$$EXP = a + b + c + (d * e * f) + g + h$$

durch den Compiler von links nach rechts



Seq. Ausführung

S1: $t1 = a + b$
 S2: $t1 = t1 + c$
 S3: $t2 = d * e$
 S4: $t2 = t2 * f$
 S5: $t1 = t1 + t2$
 S5: $t1 = t1 + g$
 S6: $t1 = t1 + h$

Parallele Ausf.

S1: $t1 = a + b$; S3: $t2 = d * e$
 S2: $t1 = t1 + c$; S4: $t2 = t2 * f$
 S5: $t1 = t1 + t2$
 S5: $t1 = t1 + g$
 S6: $t1 = t1 + h$

- Höhe:
 - längster Pfad von der Wurzel zum Blatt
 - Minimale Anzahl von Schritten zur Auswertung des Ausdrucks