

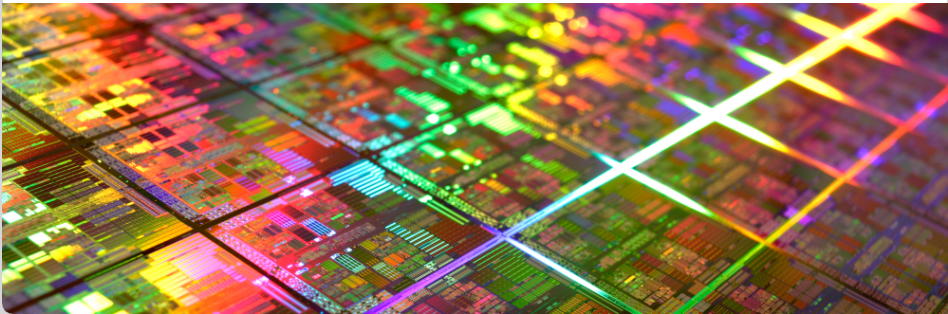
Zentralübung Rechnerstrukturen im SS 2017

Low-Power-Entwurf und Leistungsbewertung

Markus Hoffmann, Thomas Becker, Prof. Dr. Wolfgang Karl

Lehrstuhl für Rechnerarchitektur und Parallelverarbeitung

4. Mai 2017



Low-Power-Entwurf

- Leistungsverbrauch

Leistungsbewertung

- **Quantifizierung der Leistungsfähigkeit?**
 - Was ist Leistungsfähigkeit?
 - Was bedeutet "schneller"?
 - Wie sind Systeme zu vergleichen?
- **Entscheidung bei Entwurf, Auswahl und Veränderung von Rechenanlagen**
 - Objektive Quantifizierung
 - Erfassen von Teilaspekten
 - Erfassen des gesamten Systems

Leistungsverbrauch

$$P_{total} = P_{switching} + P_{shortcircuit} + P_{static} + P_{leakage}$$

$P_{switching}$: Leistungsaufnahme durch Umladen der kapazitiven Last

Schaltleistung: $P_{switching} = C_{eff} * U^2 * f$

$P_{shortcircuit}$: Leistungsaufnahme aufgrund von Kurzschluss im CMOS-Gatter bei Zustandsänderung

P_{static} : Statische Leistungsaufnahme der Schaltung

$P_{leakage}$: Leckströme

Leckströme

- Einfluss des Leckstroms steigt durch zunehmende Integrationsdichte
- Temperaturerhöhung steigert Einfluss der Leckstroms
- Variation der Versorgungsspannung und Taktfrequenz

$$P \sim f * U^2$$

- Je geringer die Versorgungsspannung, desto geringer die maximal mögliche Taktfrequenz

$$f \sim U$$

⇒ **Kubus-Regel** für simultane Änderung der Versorgungsspannung und Taktfrequenz:

$$P \sim U^3 \qquad P \sim f^3$$

Die Kernspannung von Prozessoren ist seit den 80er Jahren von 5 V auf 0,8 V gesenkt worden. Im gleichen Zeitraum stieg die Frequenz von 1 MHz auf 3 GHz.

Was bedeutet dies für die aufgenommene elektrische Leistung?

- Spannungsabsenkung:

$$U: \quad 0,8 \text{ V} \leftrightarrow 5 \text{ V}$$

$$U^2: \quad 0,64/25 = 0,0256$$

- Frequenzerhöhung:

$$f: \quad 3000 \text{ MHz}/1 \text{ MHz} = 3000$$

- Aus $P \sim U^2 * f$ resultiert eine Zunahme der elektrischen Leistung um den Faktor

$$0,0256 * 3000 = 76,8$$

Die Kernspannung von Prozessoren ist seit den 80er Jahren von 5 V auf 0,8 V gesenkt worden. Im gleichen Zeitraum stieg die Frequenz von 1 MHz auf 3 GHz.

Was bedeutet dies für die aufgenommene elektrische Leistung?
Ein neuerer Prozessor hat bei einer Frequenz von 3,5 GHz noch 0,6 V Kernspannung. Wie ändert sich dadurch die aufgenommene elektrische Leistung?

- Spannungsabsenkung:

$$U : \quad 0,6 \text{ V} \leftrightarrow 0,8 \text{ V}$$

$$U^2 : \quad 0,36 / 0,64 = 0,5625$$

- Frequenzerhöhung:

$$f : \quad 3,5 \text{ GHz} / 3 \text{ GHz} \approx 1,167$$

- $P \sim U^2 * f = 0,5625 * 1,167 \approx 0,66$

bedeutet eine Abnahme der notwendigen elektrischen Leistung.

Zum Übertakten von Prozessoren wird die Kernspannung erhöht.
Warum ist dies so?

Wie fließt die Erhöhung der Kernspannung in die Leistungsaufnahme ein und was bedeutet dies?

- Steilere Taktflanken nötig um schneller gültiges Signallevel zu erreichen
- Durch höhere Spannung schnelleres Laden von C_{eff} und damit steilere Taktflanken
- $P_{switching} = C_{eff} * U^2 * f$
- Nachteil: Spannung fließt quadratisch in Formel ein

Welcher Bestandteil der Leistungsaufnahme war früher vernachlässigbar, spielt heute jedoch eine überaus zentrale Rolle?

- Aufgrund der immer weiter ansteigenden Integrationsdichte spielen mittlerweile die **Leckströme** eine erhebliche Rolle bei der Leistungsaufnahme.
- Wegen der Leckströme führt eine Verkleinerung der Strukturen nicht automatisch zu einer Reduzierung der Stromaufnahme, was eine höhere Taktung ermöglicht.
- Leckströme steigen mit höherer Temperatur.

Schaltwahrscheinlichkeiten sind im Low-Power-Bereich von Bedeutung.

Signalwahrscheinlichkeit - Beispiel: UND-Gatter

Gegeben sei ein UND-Gatter mit zwei Eingängen. Die Eingangswerte 0, 1 seien gleichverteilt.

- UND: 1, wenn beide Eingänge 1, sonst 0
- 4 Möglichkeiten (00, 01, 10, 11), nur 11 ergibt 1 am Ausgang

- **Signalwahrscheinlichkeit:**

$$\mathbb{P}_{\text{Ausgang}}(1) = \frac{1}{4}, \quad \mathbb{P}_{\text{Ausgang}}(0) = \frac{3}{4}$$

- Berechnung auch über boolesche Funktion möglich:

$$\begin{aligned}\mathbb{P}_{\text{Ausgang}}(1) &= \mathbb{P}(a = 1 \wedge b = 1) = \mathbb{P}(a = 1) * \mathbb{P}(b = 1) \\ &= \frac{1}{2} * \frac{1}{2} = \frac{1}{4}\end{aligned}$$

Schaltwahrscheinlichkeit - Allgemeine Formel

Wahrscheinlichkeit, dass Gatter schaltet:

$$\begin{aligned}\mathbb{P}_{Schalt} &= \mathbb{P}(0 \rightarrow 1 \vee 1 \rightarrow 0) \\ &= \mathbb{P}(0 \rightarrow 1) + \mathbb{P}(1 \rightarrow 0) \\ &= \mathbb{P}(0) * \mathbb{P}_{neu}(1) + \mathbb{P}(1) * \mathbb{P}_{neu}(0) \\ &= \mathbb{P}(0) * \mathbb{P}(1) + \mathbb{P}(0) * \mathbb{P}(1) \\ &= 2 * \mathbb{P}(1) * \mathbb{P}(0) \\ &= 2 * \mathbb{P}(1) * (1 - \mathbb{P}(1))\end{aligned}$$

Daher zunächst Berechnung von $\mathbb{P}(1)$ je Gatter

Zur Ermittlung der Schaltwahrscheinlichkeit einer Schaltung wird häufig ein statistisches Modell herangezogen. Geben Sie eine allgemeine Formel zur Berechnung der Schaltwahrscheinlichkeit $\mathbb{P}_{Schalt}$ an und berechnen Sie diese für ein ODER-Gatter mit $\mathbb{P}_{Eingang\ 1=1} = \frac{1}{4}$ und $\mathbb{P}_{Eingang\ 2=1} = \frac{3}{4}$.

Allgemeine Formel

$$\mathbb{P}_{Schalt} = 2 * \mathbb{P}(1) * (1 - \mathbb{P}(1))$$

Allgemeine Formel der Schaltwahrscheinlichkeit

$$\mathbb{P}_{Schalt} = 2 * \mathbb{P}(1) * (1 - \mathbb{P}(1))$$

- Signalwahrscheinlichkeit ODER-Gatter:

$$\begin{aligned}\mathbb{P}_{Ausgang}(1) &= 1 - \mathbb{P}_{Ausgang}(0) \\ &= 1 - \mathbb{P}(a = 0 \wedge b = 0) \\ &= 1 - (1 - \frac{1}{4}) * (1 - \frac{3}{4}) = 1 - \frac{3}{4} * \frac{1}{4} = \frac{13}{16}\end{aligned}$$

- Alternative (direkte) Berechnung:

$$\begin{aligned}\mathbb{P}_{Ausgang}(1) &= \mathbb{P}(a = 1 \wedge b = 0) + \mathbb{P}(a = 0 \wedge b = 1) \\ &\quad + \mathbb{P}(a = 1 \wedge b = 1) \\ &= \frac{1}{4} * \frac{1}{4} + \frac{3}{4} * \frac{3}{4} + \frac{1}{4} * \frac{3}{4} \\ &= \frac{1}{16} + \frac{9}{16} + \frac{3}{16} = \frac{13}{16}\end{aligned}$$

Allgemeine Formel der Schaltwahrscheinlichkeit

$$\mathbb{P}_{Schalt} = 2 * \mathbb{P}(1) * (1 - \mathbb{P}(1))$$

- Signalwahrscheinlichkeit ODER-Gatter:

$$\mathbb{P}_{Ausgang}(1) = \frac{13}{16}$$

- Schaltwahrscheinlichkeit ODER-Gatter:

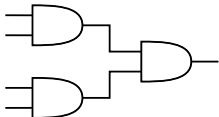
$$\mathbb{P}_{Schalt} = 2 * \mathbb{P}(1) * (1 - \mathbb{P}(1))$$

$$\begin{aligned}\mathbb{P}_{Schalt} &= 2 * \frac{13}{16} * (1 - \frac{13}{16}) \\ &= \frac{2*13*3}{16*16} = \frac{39}{128}\end{aligned}$$

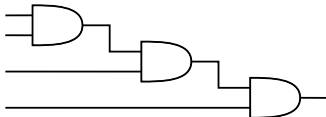
Auswirkung von Schaltwahrscheinlichkeiten:

- Inwiefern unterscheiden sich die zwei Implementierungen hinsichtlich ihres Schaltverhaltens und dem damit verbundenen Leistungsverbrauch?
- $\mathbb{P}_{\text{Eingangssignal}=1} = \mathbb{P}_{\text{Eingangssignal}=0} = 0,5$

Variante 1:



Variante 2:



Variante 1:

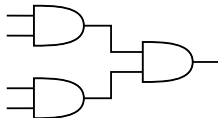
- Beide linken UND-Gatter:

Signalwahrscheinlichkeit:

$$\mathbb{P}_{links}(1) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

Schaltwahrscheinlichkeit:

$$\mathbb{P}_{Schalt-links} = 2 * \frac{1}{4} * \frac{3}{4} = \frac{3}{8}$$



- Rechtes UND-Gatter:

Signalwahrscheinlichkeiten für Eingänge des rechten Gatters =
Ausgangssignalwahrscheinlichkeiten der linken Gatter

$$\mathbb{P}_{rechts}(1) = \frac{1}{4} * \frac{1}{4} = \frac{1}{16}$$

$$\mathbb{P}_{Schalt-rechts} = 2 * \frac{1}{16} * \frac{15}{16} = \frac{15}{128}$$

- **Summe** Schaltwahrscheinlichkeiten = $\frac{3}{8} + \frac{3}{8} + \frac{15}{128} = \frac{111}{128}$

Variante 2:

- Signalwahrscheinlichkeit:

$$\mathbb{P}_{links}(1) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

Schaltwahrscheinlichkeit:

$$\mathbb{P}_{Schalt-links} = 2 * \frac{1}{4} * \frac{3}{4} = \frac{3}{8}$$

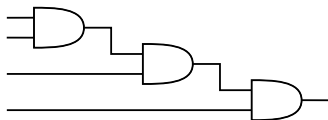
- $\mathbb{P}_{mitte}(1) = \frac{1}{2} * \frac{1}{4} = \frac{1}{8}$

$$\mathbb{P}_{Schalt-mitte} = 2 * \frac{1}{8} * \frac{7}{8} = \frac{7}{32}$$

- $\mathbb{P}_{rechts}(1) = \frac{1}{2} * \frac{1}{8} = \frac{1}{16}$

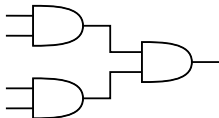
$$\mathbb{P}_{Schalt-rechts} = 2 * \frac{1}{16} * \frac{15}{16} = \frac{15}{128}$$

- $\text{Summe}_{Schaltwahrscheinlichkeiten} = \frac{3}{8} + \frac{7}{32} + \frac{15}{128} = \frac{91}{128}$



Auswirkung von Schaltwahrscheinlichkeiten:

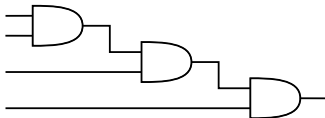
Variante 1:



■ $Summe_{Schaltw'keiten} = \frac{111}{128}$

- Höherer Leistungsverbrauch
- Geringere Durchlaufzeit

Variante 2:



■ $Summe_{Schaltw'keiten} = \frac{91}{128}$

- Geringerer Leistungsverbrauch
- Höhere Durchlaufzeit

Was ist Leistung?

■ **Anwendersicht:** Reduzierung von

- Antwortzeit (response time)
 - Latenzzeit
 - CPU Time (User, System)
- Ausführungszeit (execution time)

■ **Betriebssicht:** Erhöhung von

- Anzahl durchgeführter Jobs
- Durchsatz
- Energieeffizienz (Betriebskosten)

⇒ **Auswertung benötigt Bewertungsverfahren**

■ Auswertung von HW-Eigenschaften

- Einfacher Vergleich
- Bewertung sehr spezieller Aspekte (Takt)
- Angabe einer hypothetischen Maximalleistung (MIPS)
- Meist nicht/selten aussagekräftig
- Alltagsbeispiel: GHz-Manie → QuantiSpeed

■ Mixe

- Theoretische Berechnung einer mittleren Operationszeit T aus den Operationszeiten und Auftrethäufigkeiten von n Befehlen
- $T = \sum_{i=1}^n p_i * t_i$ mit $\sum_{i=1}^n p_i = 1$ wobei $p_i \leq 1$

■ Kernprogramme

- Typische Anwenderprogramme, für den zu bewertenden Rechner geschrieben
- Berechnung der Ausführungszeit anhand der Ausführungszeiten der Befehle

■ Benchmarks

- Programmsammlungen im Quellcode
- Übersetzung & Messung der Ausführungszeiten
- Problem: Einfluss von OS und Compiler
- Synthetische Benchmarks (Whetstone, Dhrystone), Quasi-Simulation von Anwenderprogrammen
- Kernels (LINPACK)
- Standardisierte Benchmarks (SPEC, TPC, EEMBC...)

■ Ziele

- Stellt fairen Vergleich sicher
- Ermöglicht Angabe einer Maximalleistung

■ SPEC-Benchmark

- Integer vs. Fließkomma: SPECint / SPECfp
- Geschwindigkeit vs. Durchsatz (rate)
- Optimierung: Konservativ (base) vs. aggressiv
- $SPECratio_x = \frac{t_{ref_x}}{t_{exec}}$

■ Monitore

- Gezielte Abfrage und Akkumulation von HW-Ereignissen
- Software-Monitore
- Werkzeug zur Optimierung, weniger zur Klassifizierung
- Nachteil: Beeinträchtigung des Systemverhaltens (SW)

■ Analytische Methoden

- Deterministisch (feste Werte)
- Stochastisch (verwendet bestimmte Verteilung)
- Operationell (gemessen in festem Zeitintervall)

■ Simulationen

- Modellbildung
- Deterministische, stochastische oder aufzeichnungsgesteuerte Simulation

- **Prozessortakt** gibt lediglich den Arbeitstakt (min/typ/max) des Prozessors an.
 - Kein Maß für Leistungsfähigkeit, da keine Aussage über Effizienz, Güte des Befehlssatzes etc.
 - Beispiele: Pentium4 vs. Pentium-M
- **CPI** ist ein Maß für die Effizienz einer Architektur.
 - Unterschied zwischen maximalen CPI unter Idealbedingungen und realen, programmabhängig gemessenen CPI
 - Zur Leistungsbewertung als **alleinige** Maßzahl nicht ausreichend: Effizienz $\neq$ Geschwindigkeit!
- **MIPS** auf den ersten Blick ideal, weil zwei Maßzahlen (Takt, CPI) zusammengeführt werden.
 - Aufgrund des CPI-Einflusses jedoch ebenfalls vom ausgeführten Programm abhängig
 - Nur unter gleichen Bedingungen (Sourcecode, Compiler, OS) direkt vergleichbar.

- **CPI** (Zyklen pro Instruktion)

$$CPI = \frac{c}{i}$$

- **MIPS** (Million Instructions per Second)

$$MIPS = \frac{i}{t * 10^6} = \frac{f}{CPI * 10^6}$$

- **Taktrate** (Frequenz)

$$f = \frac{c}{t} = \frac{i * CPI}{t} \text{ [Hz]}$$

- **CPU-Zeit**

$$t_{cpu} = c * t_{Zyklus}$$

- **Beachten Sie auch mögliche Umformungen!**

Welche wichtigen Architekturparameter beeinflussen jeweils die Zykluszeit, die Anzahl der Instruktionen und den CPI-Wert?

- Die **Zykluszeit** hängt von der Organisation und der Technologie ab.
- Die **Anzahl der Instruktionen** ist bedingt durch die Befehlssatzarchitektur und die Güte des Compilers.
- Die **Zyklen pro Instruktion** werden durch die Organisation und die Befehlssatzarchitektur beeinflusst.

Leistungsbewertung – Aufgabe 2

Prozessor A arbeitet ein Problem in 2 ms ab. Er hat ein CPI von 7/5 und benötigt 3.500.000 Instruktionen für die Abarbeitung der Problemstellung. Prozessor B arbeitet dieses Problem ebenfalls in 2 ms ab. Er hat ein CPI von 3/2 und benötigt 1.500.000 Instruktionen.

- **Welcher Prozessor ist für dieses Problem zu wählen und warum?**

$$f = \frac{i * CPI}{t}, \quad MIPS = \frac{f}{CPI * 10^6}$$

$$f_A = \frac{3,5 * 10^6 * \frac{7}{5}}{2 * 10^{-3} s} = 2450 MHz, \quad MIPS_A = \frac{2,45 * 10^9}{\frac{7}{5} * 10^6 s} = 1750 MIPS$$

$$f_B = \frac{1,5 * 10^6 * \frac{3}{2}}{2 * 10^{-3} s} = 1125 MHz, \quad MIPS_B = \frac{1,125 * 10^9}{\frac{3}{2} * 10^6 s} = 750 MIPS$$

$$i_A = 3.500.000, \text{ CPI}_A = \frac{7}{5}$$

$$f_A = 2450 \text{ MHz}, \text{ MIPS}_A = 1750 \text{ MIPS}$$

$$i_B = 1.500.000, \text{ CPI}_B = \frac{3}{2}$$

$$f_B = 1125 \text{ MHz}, \text{ MIPS}_B = 750 \text{ MIPS}$$

$$t = 2 \text{ ms}$$

■ Welcher Prozessor ist für dieses Problem zu wählen und warum?

■ Prozessor B, weil

- ohne Berechnung: Gleich schnell in der Abarbeitung bei wesentlich weniger Instruktionen (1,5 vs. 3,5 Mio Instruktionen)
- halbe Taktfrequenz ($P \sim U^2 * f$, Fertigung)

Leistungsbewertung – Aufgabe 3

Benchmarks sind eine verlässliche Methode zur Leistungsbewertung. Auf einem 4 GHz-Prozessor wird ein solcher Benchmark abgearbeitet. Nachfolgende Tabelle listet die auftretenden Befehlstypen mit Häufigkeit und jeweiliger Zyklenzahl.

Befehlstyp	Anzahl in 10^3	Zyklenzahl
Integer-Arithmetik	300	1
Fließkomma-Arithmetik	75	2
Speicherzugriff	150	3
Kontrollflusstransfer	25	4

Zu bestimmen sind die Werte für Ausführungszeit, CPI, MIPS und MFLOPS.

Befehlstyp	Anzahl in 10^3	Zyklenzahl
Integer-Arithmetik	300	1
Fließkomma-Arithmetik	75	2
Speicherzugriff	150	3
Kontrollflusstransfer	25	4

■ Anzahl Instruktionen

$$i = \sum i_{\text{typ}} = (300 + 75 + 150 + 25) * 10^3 = 550.000$$

■ Taktzyklen

$$\begin{aligned} c &= \sum i_{\text{typ}} * c_{\text{typ}} \\ &= (300 * 1 + 75 * 2 + 150 * 3 + 25 * 4) * 10^3 = 1.000.000 \end{aligned}$$

■ Zykluszeit bei 4 GHz Taktfrequenz

$$t = \frac{1}{f} = \frac{1}{4\text{GHz}} = 0,25 * 10^{-9} \text{s} = 0,25 \text{ ns}$$

■ Angaben und bisherige Berechnungen:

$$f = 4 \text{ GHz} \rightarrow t_c = 0,25 \text{ ns}$$

$$i = 550.000, c = 1.000.000 = 1 * 10^6$$

■ Ausführungszeit

$$t_{\text{exec}} = c * t_{\text{cyc}}$$

$$= 1 * 10^6 * 0,25 * 10^{-9} \text{ s} = 250 * 10^{-6} \text{ s} = 250 \mu\text{s}$$

■ CPI

$$\text{CPI} = \frac{c}{i} = \frac{1 * 10^6}{550 * 10^3} = \frac{100}{55} = \frac{20}{11} \approx 1,82$$

■ MIPS

$$\text{MIPS} = \frac{i}{t * 10^6} = \frac{550.000}{250} = 2200$$

■ MFLOPS

- Wie MIPS, wobei Anzahl der Befehle und Ausführungszeit nur für Fließkommaberechnung

$$\text{MFLOPS} = \frac{75.000}{(75.000 * 2) * (0,25 * 10^{-9}) * 10^6} = \frac{1}{0,5 * 10^{-3}} = 2000$$

Leistungsbewertung – Aufgabe 4

(vergl. Hennessy and Patterson, Computer Architecture A Quantitative Approach, 4. Auflage, S. 43-44.)

Sie haben für Ihre neue Rechnerarchitektur die folgenden Werte experimentell bestimmt:

Befehlstyp	CPI	Anteil
Fließkomma-Arithmetik	4,0	25 %
Restliche Befehle	1,33	75 %

Die Häufigkeit der Instruktion FPSQR beträgt 2 % und der $CPI_{FPSQR} = 20$. Es gibt zwei Entwurfsmöglichkeiten:

- (a) senken des CPI_{FPSQR} auf 2
- (b) senken des CPI-Wert der Gleitkommaoperationen auf 2,5.

Berechnen Sie den jeweiligen Gesamtgewinn der Alternativen und begründen Sie die Entscheidung.

Leistungsbewertung – Aufgabe 4

Es ändern sich nur die Zyklen pro Instruktion, Taktrate und Anzahl der Instruktionen (i) bleiben gleich.

Der unoptimierte CPI-Wert errechnet sich nach:

$$CPI_{\text{base}} = \sum_{i=1}^n CPI_i * Anteil_i = (1,33 * 75\%) + (4 * 25\%) \approx 2,0$$

Die Zyklen pro Instruktion mit neuem FPSQR-Befehl: $CPI_{(a)}$ kann durch Abziehen der gesparten Zyklen erfolgen:

$$\begin{aligned} CPI_{(a)} &= CPI_{\text{base}} - 0,02 * (CPI_{\text{old FPSQR}} - CPI_{\text{new FPSQR}}) \\ &= 2,0 - 0,02 * (20 - 2) = 1,64 \end{aligned}$$

Alternative (b) errechnet sich analog zum CPI_{base} :

$$CPI_{(b)} = (1,33 * 75\%) + (2,5 * 25\%) = 1,625$$

Aufgrund des **geringeren** CPI-Werts bietet sich die **Alternative (b)** mit den verbesserten Zyklen pro Gleitkommaoperation an.

Leistungsbewertung – Aufgabe 4

Berechnung des Gewinns (Speedup) durch die Verwendung der Alternative (*b*) gegenüber dem vorherigen System (*base*):

$$\begin{aligned} \text{Speedup}_{(b)} &= \frac{\text{CPU time}_{base}}{\text{CPU time}_{(b)}} \\ &= \frac{i * \text{Taktrate} * \text{CPI}_{base}}{i * \text{Taktrate} * \text{CPI}_{(b)}} \\ &= \frac{\text{CPI}_{base}}{\text{CPI}_{(b)}} \end{aligned}$$

Eingesetzt ergibt sich:

$$\text{Speedup}_{(b)} = \frac{2,00}{1,62} \approx 1,23$$

→ Alternative (*b*) ist 1,23-mal schneller als bisheriges System.

Die Ergebnistabelle der SPEC-Seite für die Xeon X5677-Architektur gliedert sich in die Spalten **Base** und **Peak**. Erklären Sie den Laufzeitunterschied für 400.perlbench. Vergleichen Sie dies mit den Ergebnissen für 483.xalanbmk.

- **Peak** erlaubt aggressive Optimierungen im Gegensatz zu Base. → Laufzeitunterschied durch Optimierung
 - Kaum Laufzeitunterschiede für 483.xalanbmk:
 - Entweder waren die durchgeführten Optimierungen nicht wirkungsvoll,
 - oder weitere Optimierungen wurden nicht angestrebt.
- ⇒ Sektion **Peak Optimization Flags** zeigt, dass Compileroptimierungen nicht verwendet wurden.

Leistungsbewertung – Aufgabe 5b)

Berechnen Sie unter Zuhilfenahme des Formelwerks aus der Vorlesung die **Referenzzeit** für den 462.libquantum Benchmark.

Es gilt $SPEC_{ratio} = \frac{Referenzzeit_x}{Laufzeit_x \text{ auf Testsystem}}$ für einen Benchmark x.

Die Tabelle auf der angegebenen Webseite enthält die **Laufzeiten der Benchmarks** auf dem Testsystem und die $SPEC_{ratio}$.

Leistungsbewertung – Aufgabe 5b)

Results Table				
Benchmark	Base		Peak	
	Seconds	Ratio	Seconds	Ratio
400.perlbench	344	28.4	291	33.6
401.bzip2	443	21.8	442	21.8
403.gcc	303	26.6	269	29.9
429.mcf	198	46.0	174	52.3
445.gobmk	389	26.9	358	29.3
456.hmmer	177	52.8	172	54.4
458.sjeng	426	28.4	408	29.7
462.libquantum	33.8	613	33.8	613
464.h264ref	529	41.8	486	45.5
471.omnetpp	271	23.1	214	29.2
473.astar	310	22.6	297	23.6
483.xalancbmk	168	41.1	168	41.1

Quelle: <http://www.spec.org/cpu2006/results/res2010q2/cpu2006-20100329-10254.html>

Leistungsbewertung – Aufgabe 5b)

Es gilt $SPEC_{ratio} = \frac{Referenzzeit_x}{Laufzeit_x \text{ auf Testsystem}}$

Somit ergibt sich nach dem Umstellen und Einsetzen:

$$Referenzzeit_{462.libquantum} = 613 * 33,8 \text{ s} = 20719,4 \text{ s}$$

Leistungsbewertung – Aufgabe 5c)

Welches der unter

<http://www.spec.org/cpu2006/results/cpu2006.html>

aufgeführten Systeme entspricht am ehesten dem Referenzsystem?

Die Suche ergibt **Ultra Enterprise 2** von Sun Microsystems.

Begründung:

- Die unter b) errechnete Referenzlaufzeit für den ausgewählten Benchmark stimmt annähernd überein:

Benchmark	$Referenzzeit_{errechnet}$	$Laufzeit_{Ultra\ Enterprise\ 2}$
462.libquantum	20719,4	20704

- Es wird der $SPEC_{int_base}2006 = 1.00$ angegeben.

Leistungsbewertung – Aufgabe 6

Für eine Rechenanlage soll ein geeigneter Plattenspeicher angeschafft werden. Mithilfe eines Warteschlangenmodells sollen hierzu der **Durchsatz D** und die **Auslastung U** der Plattensysteme berechnet werden unter der Annahme, die durchschnittliche **Ankunftsrate A** von Schreib-/Leseaufträgen im System liegt bei 40/s.

Zur Auswahl stehen Festplatten mit folgenden Daten:

- Platte 1: Zugriffszeit 12 ms, Datenrate 6 MByte/s
- Platte 2: Zugriffszeit 10 ms, Datenrate 7,5 MByte/s
- Platte 3: Zugriffszeit 8 ms, Datenrate 8 MByte/s

- Platte 1: Zugriffszeit 12 ms, Datenrate 6 MByte/s
- Platte 2: Zugriffszeit 10 ms, Datenrate 7,5 MByte/s
- Platte 3: Zugriffszeit 8 ms, Datenrate 8 MByte/s

Berechnen Sie für die drei Festplatten die Bedienzeit X_i , wenn der Schreib-/Leseauftrag im Schnitt 100 kB groß ist.

- **Bedienzeiten:** $X_i = t_{\text{Zugriff}} + t_{\text{Übertragung}}$

$$X_1 = 12 \text{ ms} + \frac{100 \text{ kB}}{6000 \text{ kB/s}} = 28,67 \text{ ms}$$

$$X_2 = 10 \text{ ms} + \frac{100 \text{ kB}}{7500 \text{ kB/s}} = 23,33 \text{ ms}$$

$$X_3 = 8 \text{ ms} + \frac{100 \text{ kB}}{8000 \text{ kB/s}} = 20,5 \text{ ms}$$

Leistungsbewertung – Aufgabe 6b)

Wie groß sind die Durchsätze D_i der einzelnen Festplatten?
Welche Festplatten wären aufgrund der Berechnung im System einsetzbar?

■ **Maximaler Durchsatz:** $D_{imax} = \frac{1}{X_i}$

$$D_{1max} = \frac{1}{28,67 \text{ ms}} = 34,88/\text{s}$$

$$D_{2max} = \frac{1}{23,33 \text{ ms}} = 42,86/\text{s}$$

$$D_{3max} = \frac{1}{20,5 \text{ ms}} = 48,78/\text{s}$$

Nur Platten mit $D_{max} > A$ können eingesetzt werden, da sonst die Festplatte nicht genügend Zeit hat, um alle Aufträge rechtzeitig zu bedienen.

Aufgrund von $A = 40/\text{s}$ sind nur die Platten 2 und 3 einsetzbar.

Wie groß ist die Auslastung der einsetzbaren Festplatten?

■ **Auslastung:** $U_i = D/D_{imax} = D * X_i$, hier $D = A$

$$U_2 = D * X_2 = 40/s * 23,33 ms = 0,93$$

d.h. 93% Auslastung

$$U_3 = D * X_3 = 40/s * 20,5 ms = 0,82$$

d.h. 82% Auslastung

Das Betriebssystem stelle eine FIFO-basierte Warteschlange zur Verfügung. Mit einem Monitor wurden im Betrieb hierzu ermittelt, dass die Warteschlange Q_2 von Festplatte 2 drei Aufträge umfasst, Q_3 von Festplatte 3 fasse zwei Aufträge. Berechnen Sie die Zeit der Aufträge in der Warteschlange und die Reaktionszeit des Gesamtsystems aus Warteschlange und Festplatte.

■ Gesetz von Little: $Q = W * D$

Q: Anzahl von Aufträgen in der Warteschlange

W: Wartezeit

D: Durchsatz

■ Gesetz von Little: $Q = W * D$

Q: Anzahl von Aufträgen in der Warteschlange

W: Wartezeit

D: Durchsatz

d.h. $W_i = \frac{Q_i}{D}$, wobei abermals gilt $D = A$ und somit

$$W_2 = \frac{Q_2}{D} = \frac{3}{40/s} = 75 \text{ ms}$$

$$W_3 = \frac{Q_3}{D} = \frac{2}{40/s} = 50 \text{ ms}$$

- Reaktionszeit des Gesamtsystems aus Warteschlange und Festplatte: $Reaktionszeit_i = Wartezeit_i + Bedienzeit_i$
- bereits berechnet: $W_2 = 75\ ms$, $W_3 = 50\ ms$ und $X_2 = 23,33\ ms$, $X_3 = 20,5\ ms$
- einsetzen ergibt:
 $Reaktionszeit_2 = 75\ ms + 23,33\ ms = 98,33\ ms$
 $Reaktionszeit_3 = 50\ ms + 20,5\ ms = 70,5\ ms$

Damit ist das System mit Platte 3 vorzuziehen, da es schneller reagiert.

Fragen?

Zentralübung Rechnerstrukturen im SS 2017

Low-Power-Entwurf und Leistungsbewertung

Markus Hoffmann, Thomas Becker, Prof. Dr. Wolfgang Karl

Lehrstuhl für Rechnerarchitektur und Parallelverarbeitung

4. Mai 2017

