

AFIPS Frühjahr gemeinsame Computerkonferenz, 1967

Gültigkeit des Einzelprozessor-Ansatzes zur Erzielung eines großen Maßstabs Rechenfähigkeiten ¹

Gene M. Amdahl
IBM Sunnyvale, Kalifornien

1. EINLEITUNG

Seit über einem Jahrzehnt äußern Propheten die Behauptung, dass die Organisation eines einzelnen Computers hat seine Grenzen erreicht und dass wirklich bedeutende Fortschritte nur durch die Verbindung von a gemacht werden können Vielzahl von Computern so, dass eine kooperative Lösung möglich ist. Verschiedenes das Richtige Richtung wurde als Allzweckcomputer mit einer verallgemeinerten Verbindung herausgestellt von Speichern oder als spezialisierte Computer mit geometrisch verwandten Speicherverbindungen und gesteuert durch einen oder mehrere Befehlsströme.

Es wird die fortgesetzte Gültigkeit des Einzelprozessor-Ansatzes und des Schwächen des Mehrprozessor-Ansatzes in Bezug auf die Anwendung auf reale Probleme und deren begleitende Unregelmäßigkeiten.

Die vorgebrachten Argumente basieren auf statistischen Merkmalen der Berechnung auf Computern im letzten Jahrzehnt und auf die betrieblichen Anforderungen innerhalb von Problemen von physischem Interesse. Eine zusätzliche Referenz wird eine der gründlichsten Analysen der relativen Computerfähigkeiten sein Derzeit veröffentlicht \ Änderungen in der Computerleistung. " Datamation , September 1966, Professor Kenneth F. Knight von der Stanford School of Business Administration.

Das erste interessierende Merkmal ist der Anteil der damit verbundenen Rechenlast mit Datenverwaltung Housekeeping. Dieser Anteil ist seit etwa zehn Jahren nahezu konstant Jahre und macht 40% der ausgeführten Anweisungen in Produktionsläufen aus. In einem ganz engagierten In einer speziellen Umgebung kann dies um den Faktor zwei reduziert werden, dies ist jedoch höchst unwahrscheinlich dass es um den Faktor drei reduziert werden könnte. Die Art dieses Overheads scheint sequentiell zu sein so dass es unwahrscheinlich ist, dass es parallelen Verarbeitungstechniken zugänglich ist. Overhead allein würde dann Legen Sie eine Obergrenze für den Durchsatz fest, die das Fünf- bis Siebenfache der sequentiellen Verarbeitungsrates beträgt, selbst wenn die

¹ Dieses Papier wurde von Guihai Chen als vorliegende Form neu getippt. Er wünscht, Sie würden gerne dieses historische lesen Papier.

Die Reinigung erfolgte in einem separaten Prozessor. Der nicht haushaltsübliche Teil des Problems könnte Nutzen Sie höchstens einen Prozessor mit der drei- bis vierfachen Leistung des Housekeeping Prozessor. Eine ziemlich offensichtliche Schlussfolgerung, die an dieser Stelle gezogen werden kann, ist, dass der Aufwand aufgewendet wurde Das Erreichen hoher Parallelverarbeitungsraten wird verschwendet, es sei denn, dies geht mit Erfolgen in einher sequentielle Verarbeitungsraten von nahezu der gleichen Größe.

Data Management Housekeeping ist nicht das einzige Problem, das vereinfachte Ansätze plagt Hochgeschwindigkeitsberechnung. Die körperlichen Probleme, die von praktischem Interesse sind, haben eher signifikante Komplikationen. Beispiele für diese Komplikationen sind: Grenzen sind wahrscheinlich Bei unregelmäßigen Innenräumen sind inhomogene Berechnungen erforderlich, die von den Zuständen abhängen können der Variablen an jedem Punkt können sich die Ausbreitungsraten verschiedener physikalischer Effekte stark unterscheiden Die Konvergenzrate oder Konvergenzrate kann stark davon abhängen, ob die Konvergenz durchgeführt wird Anordnung entlang verschiedener Achsen bei nachfolgenden Durchgängen usw. Die Auswirkung jeder dieser Komplikationen ist Sehr schwerwiegend bei jeder Computerorganisation, die auf geometrisch verwandten Prozessoren in einer Parallelität basiert Verarbeitungssystem. Sogar die Existenz regelmäßiger rechteckiger Grenzen hat die interessante Eigenschaft dass für die räumliche Dimension von N^3 verschiedene Punktgeometrien in einem nächsten behandelt werden müssen Nachbarberechnung. Wenn auch der zweitnächste Nachbar beteiligt wäre, gäbe es $5N$ verschiedene Punktgeometrien zu bewältigen. Eine unregelmäßige Grenze verstärkt dieses Problem ebenso wie dies ein inhomogenes Interieur. Berechnungen, die von den Zuständen der Variablen abhängen, würden erfordern, dass die Verarbeitung an jedem Punkt ungefähr die gleichen Rechenzeiten wie verbraucht die Summe der Berechnungen aller physikalischen Effekte innerhalb einer großen Region. Unterschiede der Änderungen in Ausbreitungsraten können die Netzpunktbeziehungen beeinflussen.

Idealerweise die Berechnung der Wirkung der benachbarten Punkte auf den betrachteten Punkt Bei der Berechnung werden ihre Werte zu einem früheren Zeitpunkt proportional zum Maschenabstand und umgekehrt berücksichtigt proportional zur Ausbreitungsrate. Da der Zeitschritt normalerweise konstant gehalten wird, wird eine schnellere Die Agationsrate für einige Effekte würde Wechselwirkungen mit weiter entfernten Punkten implizieren. Endlich das faire gängige Praxis, bei aufeinanderfolgenden Durchläufen entlang verschiedener Achsen durch das Netz zu fegen Probleme der Datenverwaltung, die alle Prozessoren betreffen, jedoch geometrisch verwandte Prozessoren strenger, indem zusätzlich zu den überarbeiteten Punkten alle Speicherpunkte transponiert werden müssen Eingabe-Ausgabe-Planung. Eine realistische Einschätzung der Auswirkungen dieser Unregelmäßigkeiten auf eine vereinfachte und eine regulierte Abstraktion des Problems führt zu einer Verschlechterung in der Nähe von einer Hälfte zu eins Größenordnung.

Um die Auswirkungen der Datenverwaltung und der Problemunregelmäßigkeiten zusammenzufassen, Der Autor hat drei verschiedene Maschinenorganisationen mit ungefähr gleichen Beträgen verglichen

von Hardware. Maschine A hat zweiunddreißig arithmetische Ausführungseinheiten, die durch einen einzelnen Befehl gesteuert werden Strom. Maschine B verfügt über Pipeline-Arithmetik-Ausführungseinheiten mit bis zu drei überlappenden Operationen auf Vektoren von acht Elementen. Maschine C hat die gleichen Pipeline-Ausführungseinheiten, jedoch die Initiierung von Einzeloperationen mit der gleichen Geschwindigkeit wie Maschine B erlaubten Vektorelementoperationen. Das Die Leistung dieser drei Maschinen ist in Abbildung 1 als Funktion des Anteils der dargestellt Anzahl der Anweisungen, die Parallelität zulassen. Der wahrscheinliche Betriebsbereich ist zentriert um einen Punkt, der 25% des Datenverwaltungsaufwands und 10% der Problemoperationen entspricht

gezwungen, sequentiell zu sein.

Abbildung 1

Die historische Leistung im Vergleich zu den Kosten von Computern wurde von Professor Knight. Die sorgfältig analysierten Daten, die er präsentiert, spiegeln nicht nur die Ausführungszeiten für die Arithmetik wider Betrieb und Kosten des Minimums der empfohlenen Konfigurationen. Er enthält Speicherkapazität effekte, erfahrene Überschneidungen zwischen Eingabe und Ausgabe und spezielle Funktionsfähigkeiten. Die beste statistische Anpassung Die erhaltene Leistung entspricht einer Leistung, die proportional zum Quadrat der Kosten auf jeder technologischen Ebene ist. Dieses Ergebnis unterstützt sehr effektiv das oft angeführte "Grosch'sche Gesetz". Unter Verwendung dieser Analyse eine Ich kann argumentieren, dass man erwarten könnte, wenn die doppelte Menge an Hardware in einem einzigen System ausgenutzt würde um die vierfache Leistung zu erhalten. Die einzige Schwierigkeit besteht darin, zu wissen, wie man dies ausnutzt zusätzliche Hardware. Zu jedem Zeitpunkt ist es schwierig vorherzusagen, wie sich die kostbaren Engpässe auswirken Ein sequentieller Computer wird effektiv überwunden. Wenn es einfach wäre, wären sie nicht als zurückgelassen worden Engpässe. Es ist ein historisches Beispiel, dass die aufeinanderfolgenden Hindernisse überwunden wurden, so ist es auch angemessen, um Rev. Adam Clayton Powell zu zitieren ("Behalte den Glauben, Baby!"). Wenn alternativ einer beschlossen, die Leistung zu verbessern, indem zwei Prozessoren neben den gemeinsam genutzten Speicher gestellt werden. man würde ungefähr 2,2 mal so viel Hardware finden. Die zusätzlichen zwei Zehntel in Hardware hat die Crossbar-Umschaltung für die Freigabe durchgeführt. Die daraus resultierende Leistung würde etwa 1,8 sein. Die letztere Zahl wird aus der Annahme abgeleitet, dass jeder Prozessor die Hälfte von verwendet

3

Seite 4

die Erinnerungen etwa die Hälfte der Zeit. Die daraus resultierenden Speicherkonflikte im gemeinsam genutzten System würden Verlängern Sie die Ausführung einer von zwei Operationen um ein Viertel der Ausführungszeit. Das Nettoergebnis ist eher eine Verschlechterung der Preisleistung auf 0,8 als eine Verbesserung auf 2,0 für die größere Einzelperson Prozessor.

Eine vergleichende Analyse mit einem assoziativen Prozessor ist weitaus weniger einfach und offensichtlich. Unter bestimmten Bei regulären Formaten gibt es einen ziemlich direkten Ansatz. Betrachten Sie einen assoziativen Prozessor signiert für die Mustererkennung, bei der Entscheidungen innerhalb einzelner Elemente an einige weitergeleitet werden Satz anderer Elemente. In dem assoziativen Prozessordesign hätten die empfangenden Elemente einen Satz von Quelladressen, die durch assoziative Techniken erkennen, ob sie die De- empfangen sollten oder nicht Entscheidung über das aktuell deklarierende Element. Um einen entsprechenden speziellen Zweck nicht assoziativ zu machen Prozessor würde man ein empfangendes Element und seine Quelladressen als Befehl betrachten, mit binäre Entscheidung in Registern geführt. In Anbetracht der Verwendung des Filmspeichers ein Assoziativ Der Zyklus wäre länger als ein zerstörungsfreier Lesezyklus. In solch einer realen Analogie der besondere Zweck Es ist zu erwarten, dass ein nicht assoziativer Prozessor etwa ein Viertel so viele Speicherzyklen benötigt wie der assoziative Version und nur etwa ein Sechstel der Zeit. Diese Zahlen wurden vollständig berechnet

Erkennungsaufgabe mit etwas unterschiedlichen Verhältnissen in jeder Phase. Hier ist kein pauschaler Anspruch beabsichtigt, sondern dass jede Anforderung aus beiden Ansätzen untersucht werden sollte.

2 Notizen von Guihai Chen

- Dem sehr berühmten Amdahlschen Gesetz, das wie in der folgenden Formel dargestellt ist, wird dies vorenthalten Papier. Amdahl gab jedoch nur eine wörtliche Beschreibung, die von Nachzüglern umschrieben wurde wie folgt:

$$\text{Beschleunigung} = \frac{1}{r_s + \frac{r_p}{n}}$$

wobei $r_s + r_p = 1$ ist und r_s das Verhältnis des sequentiellen Teils in einem Programm darstellt.

- Nur ein kleiner Teil dieses Papiers, genau der vierte Absatz, trägt zu den Amdahls bei Gesetz. In diesem Papier wurden auch einige andere wichtige Probleme erörtert. Zum Beispiel hatte Amdahl vorausgesehen viele negative Faktoren, die die parallele Berechnung unregelmäßiger Probleme plagen, wie z als 1) Grenzen sind wahrscheinlich unregelmäßig 2) Innenräume sind inhomogen 3) Berechnungen erforderlich kann abhängig von den Zuständen der Variablen an jedem Punkt 4) Ausbreitungsraten sein von verschiedenen physikalischen Effekten kann sehr unterschiedlich sein 5) die Konvergenzrate oder Konvergenzrate überhaupt kann stark davon abhängig sein, durch das Array entlang verschiedener Achsen zu fegen nachfolgende Pässe usw.