

Programming Real-Time Multi- Computers for Signal Processing

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY, ZURICH
for the degree of
Doctor of Technical Sciences

presented by
Urban Andreas Thoeni
Dipl. El.-Ing. ETH

born January 31, 1961
citizen of Stierva, GR and Riehen, BS

accepted on the recommendation of
Prof. Dr. H.P. Geering, examiner
Prof. Dr. A. Kündig, co-examiner

Summary

This dissertation describes a research project on automatic parallelization of computations for real-time signal processing and control. The target system has the architecture of a message-passing multi-computer. The aim of the allocation is to minimize the iteration time for a set of computations which is executed periodically in a non-overlapping manner.

The algorithms to be implemented are non-recursive and allow the static determination of the size of all data used. For the analysis of the structural properties of the algorithms a data flow graph model is used. The data flow graphs are generated from the functional language SISAL.

The data flow graph is analyzed and the communication and execution costs are determined using a newly developed communication model and an instruction execution cost table for the INMOS Transputer processor.

Subsequently, the graph nodes are clustered in linear tasks. These tasks are allocated to the processors of the system. A novel two-phase allocation algorithm is presented. In the first phase, the emphasis lies on the preservation of parallelism, on the minimization of the total communication costs, and on distributing the load evenly among the processors. In the second phase, the constraints of the real target system are considered: the limited number of processors and the limited number of serial interconnections among the processors. No predefined interconnection topology is assumed; its exact structure and the number of processors actually used are determined during the allocation process.

From the partitioned and allocated data flow graph the source code for the target system is generated in OCCAM.

Extensive references to the literature are given for the application algorithms, the partitioning process, and the static allocation procedure.

Sample allocations of algorithms show a good performance. Typically, the computation time is lowered to as little as one third of the serial execution time, using two to four processors. The results prove that data flow graphs are an excellent means of representing algorithms for generating parallel implementations.

The multi-computer used at the laboratory consisting of eight Transputers and a newly developed versatile high-speed data acquisition system are briefly characterized.

Zusammenfassung

In der vorliegenden Dissertation wird ein Forschungsprojekt über die automatische Verteilung von Berechnungen beschrieben. Die Anwendungen stammen aus dem Gebiet der Echtzeit-Signalverarbeitung und Regelungen. Der Zielrechner ist aus mehreren Prozessoren zusammengesetzt, welche durch serielle Übermittlung Daten austauschen. Durch die Allokation soll die Iterationszeit minimiert werden, welche die Berechnung eines vorgegebenen Algorithmus' benötigt. Die Berechnungen werden zyklisch, aber zeitlich nicht überlappend ausgeführt.

Die zu verteilenden Algorithmen sind nicht rekursiv und können voraus auf den Speicherplatzbedarf hin untersucht werden. Um die Struktur der Algorithmen untersuchen zu können, werden sie als Datenflussgraphen dargestellt. Diese Graphen werden mittels der Sprache SISAL erzeugt.

Diese Datenflussgraphen werden auf ihre Kommunikations- und Ausführungskosten hin untersucht. Dazu werden ein neuentwickeltes Kommunikationskostenmodell und eine Instruktionskosten-Tabelle für den Transputer der Firma INMOS verwendet.

Die Knoten des Graphen werden zu linearen Prozessen zusammengefasst und auf die Prozessoren des Systems verteilt. Zu diesem Zweck wird ein neuartiger Allokations-Algorithmus angewandt, welcher in zwei Phasen zerfällt. Zu Beginn der Allokation wird darauf geachtet, dass die vorhandene Parallelität erhalten bleibt, dass die Summe der Kommunikationskosten minimiert wird und dass die Rechenlast möglichst gleichmässig auf die Prozessoren verteilt wird. Nachher treten die vom Zielsystem vorgegebenen Einschränkungen in den Vordergrund: die beschränkte Anzahl Prozessoren und die limitierte Anzahl serieller Verbindungen zwischen den Prozessoren. Zu Beginn werden keinerlei Annahmen über die Struktur der Verbindungen zwischen den Prozessoren gemacht, sie und die genaue Anzahl der eingesetzten Prozessoren werden erst während der Allokation der Prozesse festgelegt.

Aus dem partitionierten und allozierten Datenflussgraphen wird für das Zielsystem Programmcode in der Sprache OCCAM erzeugt.

Es werden ausführliche Literaturreferenzen angegeben, vor allem zu den Anwendungsalgorithmen, zur Partitionierung und zur statischen Allokation.

Die als Beispiele verteilten Algorithmen belegen die guten Leistungen des Systems. Die Rechenzeit wird typischerweise unter Verwendung von zwei bis vier Prozessoren soweit verringert, dass sie im besten Fall nur noch ein Drittel der sequentiellen Ausführungszeit beträgt. Diese Ergebnisse zeigen, dass Datenflussgraphen eine sehr geeignete Darstellungsform für Algorithmen sind, die auf Parallelrechnern implementiert werden sollen.

Das am Institut eingesetzte Mehrprozessor-System, bestehend aus acht Transputern und einem neuentwickelten universellen Hochgeschwindigkeits-Datenerfassungs-System, wird kurz beschrieben.