

Einführung in den Compilerbau

Andreas Koch

Wintersemester 2018-2019

Inhaltsverzeichnis

1 Organisatorisches	2
1.1 Literatur	2
1.2 Aufbau der Veranstaltung	2
2 Einleitung	2
2.1 Compiler	2
2.2 Auswirkung von Compilern	3
2.3 Programmiersprachen	3
2.4 Abstraktionsebenen	3
3 Zielmaschine	4
3.1 Auswirkungen der Zielmaschine	4
3.2 Anforderungen an CPUs	4
3.3 Paralleles Rechnen	6
3.4 Compilerbau in der Lehre	6
4 Aufbau	6
4.1 Vorgehen	6
4.2 Syntaxanalyse	6
4.3 Kontextanalyse	6
4.4 Code-Erzeugung	7
5 Optimierung	7
5.1 Optimierender Compiler	7
5.2 Beispiele für Optimierung	8
6 Syntax	9
6.1 Syntax oder Grammatik	9
6.2 Kontextuelle Einschränkungen	9
6.3 Semantik	9

1 Organisatorisches

1.1 Literatur

Die Vorlesung basiert **fast vollständig** auf *Programming Language Processors in Java*¹. Auszugsweise noch weiteres Material, z. B. zum ANTLR-Parsergenerator.

Einen guten allgemeinen Überblick, aber im Detail mit anderen Schwerpunkten als diese Vorlesung, bietet *Compilers, 2. Auflage*².

1.2 Aufbau der Veranstaltung

Diese Veranstaltung ist logisch in mehrere Teile gegliedert.

Front-End³ Übersicht, ca. 3 Wochen.

- Lexing und Parsing,
- Zwischendarstellungen.

Middle-End Übersicht, ca. 2 Wochen.

- Semantische- und Kontextanalyse.

Back-End Übersicht, ca. 4 Wochen.

- Laufzeitorganisation,
- Code-Erzeugung.

Front-End-Generatoren Verwendung, ca. 2–3 Wochen.

Java Virtuelle Maschine ca. 1–2 Wochen.

2 Einleitung

2.1 Compiler

Ein Compiler ist eine Schnittstelle zwischen Mensch und Maschine. Er übersetzt von einer Programmiersprache (menschenslesbar) in eine Maschinensprache (maschinenlesbar).

Programmiersprache Gut für Menschen lesbar. Beispiele für Programmiersprachen sind Smalltalk, Java und C++.

Maschinensprache Getrimmt auf Ausführungsgeschwindigkeit, Preis (Fläche), Energieverbrauch, nur sehr selten auf leichte Programmierbarkeit.

¹von David Watt und Deryck Brown, Prentice-Hall 2000

²Von Aho, Sethi, Ullmann, Lam, Addison-Wesley 2006. Auch auf Deutsch verfügbar.

³Die ersten drei Teile der Veranstaltung richten sich an die Veranstaltungen *IMT3052* von Ivar Farup, Universität Grøvik, Norwegen; und *Vertalerbouw* von Theu Ruys, Universität Twente, Niederlande.

2.2 Auswirkung von Compilern

Compiler	Ausführungszeit	Programmgröße
GCC 3.3.6	7,5 ms	13 KB
ICC 9.0	6,5 ms	511 KB

Tabelle 1: Bildkompression auf Dothan CPU, 2 GHz

Mit den Eigenschaften entscheidet ein Compiler über die dem Nutzer zugängliche Rechenleistung. Durch gewisse Abwägungen kann ein Compiler ein Programm so kompilieren, dass es am schnellsten läuft oder dass der Resultierende Code an kleinsten ist. In Tabelle 1 ist zu erkennen, welche Tradeoffs unterschiedliche Compiler machen: der von GCC generierte Code ist zwar langsamer, aber auch kleiner, während ICC schnellen Code generiert, der dafür aber auch voluminöser ist.

2.3 Programmiersprachen

Hohe Ebene Smalltalk, Java, C++. Beispiel:

```
let
  var i : Integer;
in
  i := i + 1;
```

Mittlere Ebene Assembly

```
LOAD  R1, (i)
LOADI R2, 1
ADD   R1, R1, R2
STORE R1, (i)
```

Niedrige Ebene Maschinensprache

```
0110000100000110
0111001001000001
1011000100010010
1001000100000110
```

2.4 Abstraktionsebenen

Auf den unteren Ebenen werden die Beschreibungen immer feiner, da man näher an der Zielmaschine (Hardware) arbeitet.

Der Compiler ist dafür zuständig, Details hinzuzufügen. In den obenstehenden Codebeispielen musste der Compiler unter anderem Register wählen, in denen die Werte während dem Programmablauf zwischengespeichert werden. Außerdem musste die Adresse der Variable `i`, hinzugefügt werden, wobei hier `00000110` verwendet wurde.

Diese Details werden mithilfe von verschiedensten Algorithmen ergänzt, welche die **Programmeigenschaften Analysieren** und durch die **Synthese von Details** die Beschreibung Verfeinern.

3 Zielmaschine

3.1 Auswirkungen der Zielmaschine

Die Zielmaschine hat einen Einfluss auf die Architektur des Compilers. So basiert zum Beispiel die DLX Architektur von John Hennessy und David Patterson auf der MIPS Architektur, und ist nur leicht verändert um diese zu modernisieren. Damit ist es sehr einfach, einen Compiler zu bauen, der für diese Architektur Code generiert.

Etwas komplizierter wird es mit der *TigerSHARC* Architektur von Analog Devices. Dies ist ein Beispiel für ein DSP, also *Digital Signalling Processor*. Es gibt hier zwei *Computational Blocks*, damit parallel Rechnungen ausgeführt werden können. Diese können aber nicht kommunizieren, also muss der Compiler drauf achten, dass auf die Register nur von dem jeweiligen Block aus zugegriffen werden können. Außerdem besitzt diese Architektur separate Rechenblocks, um Adressen zu bestimmen. Ein Compiler für diese Architektur muss also wissen, wie die Architektur aufgebaut ist, um sie effizient zu nutzen.

Am problematischsten wird das aber erst bei extremen Architekturen wie der des IBM/Sony *Cell* Prozessors. Hierbei handelt es sich um eine sehr gewagte, und wie sich leider herausgestellt hat, zu komplexe Architektur. Es gibt ein PowerPC-basierten Hauptprozessor, der aber nicht sehr leistungsstark ist. Die eigentliche Arbeit kann von *Synergistic Processing Elements*, kurz SPE, ausgeführt werden. Diese sind so konstruiert, dass ein sorgfältig produzierter Instruktionsstream die Hardware maximal ausnutzen kann, mit parallelen Arithmetischen- und Speichereinheiten (siehe Abbildung 3).

3.2 Anforderungen an CPUs

Je nach Anwendungsgebiet sind die Anforderungen mehr oder weniger wichtig. CPUs werden mit Spezifikationen wie der gewünschten Rechenleistung, Datentype und Operationen, die effizient verarbeitet werden sollen (Gleitkomma, Vektoren, Matrizen, Multiplikation, *multiply-accumulate*), Speicherbandbreite, Energieeffizienz und Platzbedarf. Oft können manche Anforderungen nur durch spezialisierte Prozessoren erfüllt werden.

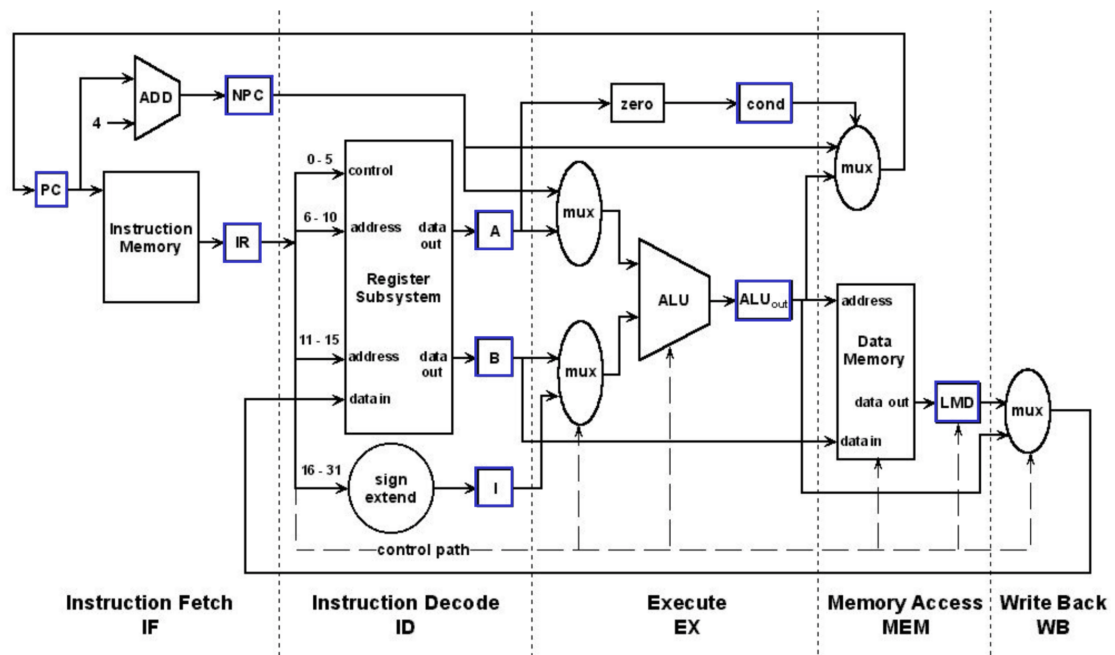


Abbildung 1: Die DLX RISC Prozessorarchitektur

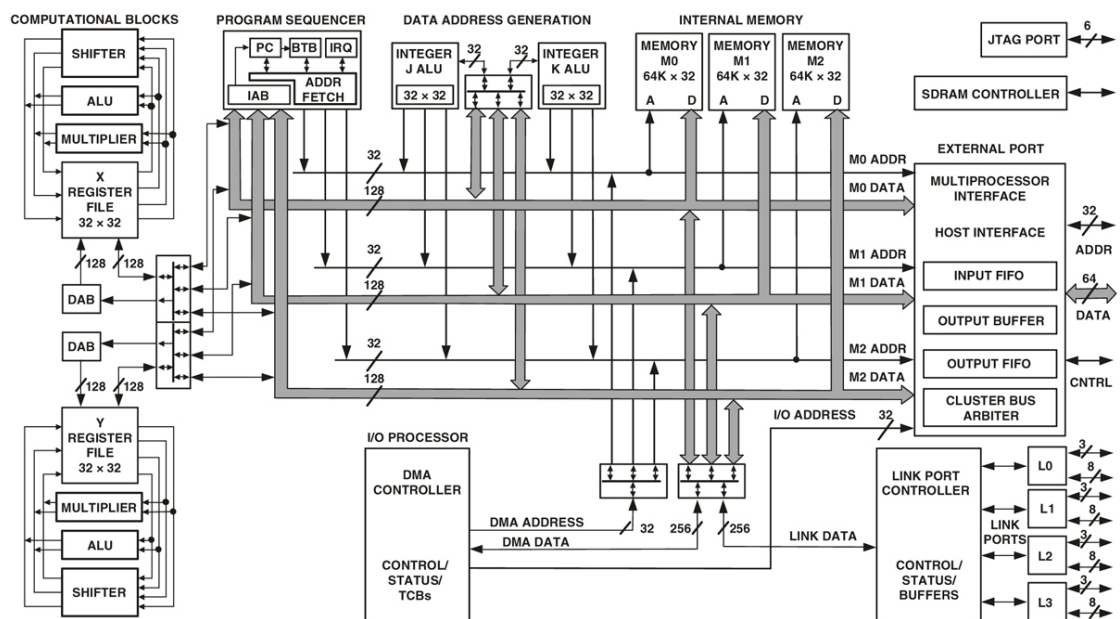


Abbildung 2: Analog Devices TigerSHARC

Abbildung 3: Ein Synergistic Processing Element eines Cell Prozessors.

3.3 Paralleles Rechnen

Paralleles Rechnen ist die Königsklasse der Forschung. Warum eigentlich? Das Wettrennen der Taktfrequenzen ist im großen und ganzen vorbei, mehr als ~4 GHz ist nicht realistisch. Also ist **der Trend weg von hochgetakteten Einzelprozessoren und hin zu vielen (2-8, teilweise 16) Prozessoren**, die aber weniger schnell getaktet werden. Dadurch erreicht man mehr Rechenleistung. Aber wie kann man für solche parallele Rechenkapazitäten programmieren?

Erste praktische Ansätze sind OpenMP, mit dem man für parallele CPUs programmieren kann. OpenCL kann mit heterogenen Systemen (GPUs, CPUs, experimentell auch FPGAs) arbeiten, und für NVidia Grafikkarten gibt es CUDA. **Diese Lösungen erfordern aber aktuell noch, dass der Programmierer explizit für ein paralleles System programmiert. Es gibt noch keine automatische Parallelisierung.**

3.4 Compilerbau in der Lehre

Warum wird der Compilerbau so früh gelehrt? Es handelt sich hierbei um eine Kombination von verschiedenen Disziplinen. Zum Compilerbau benötigt man einen *Parser*, welcher zur theoretischen Informatik gehört. Man braucht auch das Hintergrundwissen der *Architektur der Zielmaschine*, welches zur technischen Informatik gehört. Außerdem braucht man zum Compilerbau Kenntnisse von Software-Engineering, was zur praktischen Informatik gehört. Damit vereint der Compilerbau drei wichtige Disziplinen.

4 Aufbau

4.1 Vorgehen

Compiler arbeiten generell in mehreren Phasen. In Abbildung 4 sind diese Phasen aufgezeigt. Außerdem sind die Zwischendarstellungen, die zum Informationsaustausch zwischen den Phasen genutzt werden, dargestellt.

4.2 Syntaxanalyse

Bei der Syntaxanalyse wird überprüft, ob das Programm Syntaxgerecht aufgebaut ist.

4.3 Kontextanalyse

Bei der Kontextanalyse werden Variablen ihren Deklarationen zugeordnet. Außerdem werden die Typen von Ausdrücken berechnet.

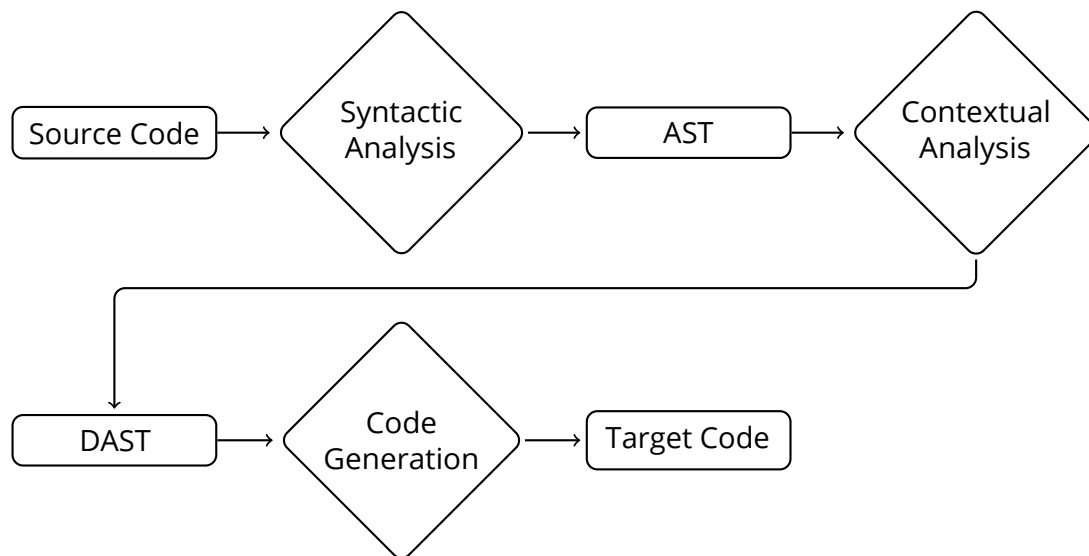


Abbildung 4: Zwischendarstellungen für den Informationsaustausch

4.4 Code-Erzeugung

Wenn das Programm syntaktisch und kontextuell korrekt ist, kann es in die Zielsprache übersetzt werden. Eine Zielsprache kann eine Low-Level Sprache wie die Maschinsprache oder Assembly, aber auch C oder eine andere Hochsprache.

Dazu wird DAST-Teilen Instruktionen der Zielsprache zugeordnet. Variablen werden gehandhabt, indem zum Beispiel bei einer Deklaration Speicherplatz reserviert wird, und bei der Verwendung der zugeordnete Speicherplatz referenziert wird.

5 Optimierung

5.1 Optimierender Compiler

Bei einem optimierenden Compiler findet die Kompilierung in drei Phasen statt.

Front-End Macht die syntaktische und kontextuelle Analyse und gibt den AST in einem IR-Format (*Intermediate Representation*) aus.

Back-End Liest das IR, und erzeugt den Code.

Middle-End Transformation von Zwischendarstellungen. Dieses arbeitet auf der IR. Hier wird kein Code ausgegeben, sondern nur der IR optimiert. Intern werden hier auch oft zusätzliche Darstellungen verwendet werden.

Anmerkung

LLVM ist ein sehr bekanntes Projekt, welches eine Art Plugin-basierte Architektur aufweist. Es können Front-Ends geschrieben werden, die unterschiedliche Programmiersprachen unterstützen (zum Beispiel `clang` für C und C++, `rubinius` für Ruby). Optimierungen können ebenso als Plugin geschrieben werden, und arbeiten auf dem LLVM IR. Es können außerdem Back-Ends geschrieben werden, die für unterschiedlichste Architekturen Code ausgeben. Somit funktioniert zum Beispiel das *Emscripten* Projekt, mit welchem man C und C++ Code in JavaScript kompilieren kann.

5.2 Beispiele für Optimierung

Constant Folding Bei dieser Optimierung werden Ausdrücke, die ein konstantes Ergebnis haben, von Optimierer ausgewertet. Hat man zum Beispiel den Code

```
int x = (2 + 3) * y;
```

Bei dem Constant Folding erkennt der Compiler, dass der Ausdruck `2 + 3` konstant ist. Er kann den Ausdruck dann bestimmen, und den Code in eine minimierte Version transformieren.

```
int x = 5 * y;
```

Anmerkung

Sowohl C++ (seit C++11) als auch Rust haben mittlerweile Support für Constant Folding in die Sprache eingebaut. Mit dem `constexpr` Keyword kann man dem Compiler sagen, dass er eine Funktion oder Variable als Konstante ansehen und diese während der Kompilierung auswerten soll.

Common Subexpression Elimination Hierbei werden Ausdrücke (*Subexpressions*), die öfters im Code vorkommt (*Common*) eliminiert. Hat man zum Beispiel folgenden Code:

```
int x = 5 * a + b;  
int y = 5 * a + c;
```

So kann der Compiler erkennen, dass der Ausdruck `5 * a` mehrmals vorkommt. Wenn er garantieren kann, dass der Wert von `a` während den Instruktionen konstant bleibt, kann er diese vereinfachen, indem er eine temporäre Variable einführt.

```
int _temp = 5 * a  
int x = _temp + b;  
int y = _temp + c;
```


6 Syntax

6.1 Syntax oder Grammatik

Der Syntax beschreibt die Satzstruktur von korrekten Programmen. Jede Sprache hat ihren eigenen Syntax. So ist zum Beispiel der folgende Ausdruck Syntaktisch korrekt in C.

```
int n = m + 1;
```

Der Satz „Ein Kreis hat zwei Ecken“ ist wiederum eine syntaktisch korrekte Aussage im Deutschen (was aber nicht bedeutet, dass diese auch *semantisch* korrekt ist).

6.2 Kontextuelle Einschränkungen

Nicht jede Aussage, die syntaktisch korrekt ist, ist in dem Kontext auch erlaubt. Es gibt gewisse Regeln, die vom Geltungsbereich (*scope*) und dem Typ von Aussagen abhängen. So ist der folgende Code zum Beispiel nur dann korrekt, wenn die Variable *x* deklariert ist.

```
int y = 2 * x;
```

Bei dem Satz „Ein Kreis hat zwei Ecken“ ist der Fehler, dass ein Kreis keine Ecken besitzt. Hier passen die Typen nicht, ein Objekt der Klasse *eckige Geometrische Formen*, wie zum Beispiel das Dreieck, würde hier kontextuell passen (aber der Satz wäre dann semantisch immer noch falsch).

6.3 Semantik