TDDC86 Compiler Optimizations and Code Generation

Instruction Selection
Retargetability

3 Main Tasks in Code Generation

- **Instruction Selection**
  - Choose set of instructions equivalent to IR code
  - Minimize (locally) execution time, # used registers, code size
  - Example: \( \text{INCRM} \#4(fp) \) vs. \( \text{LOAD} \#4(fp), R1 \text{ ADD} R1, \#1, R1 \text{ STORE} R1, \#4(fp) \)

- **Instruction Scheduling**
  - Reorder instructions to better utilize processor architecture
  - Minimize temporary space (registers, stack locations) used, execution time, or energy consumption

- **Register Allocation**
  - Keep frequently used values in registers (limited resource!)
  - Some registers are reserved, e.g. sp, fp, pc, sr, retval ...
  - Minimize #loads and #stores (which are expensive instructions!)

**Register Allocation:** Which variables to keep when in some register?

**Register Assignment:** In which particular register to keep each?

Some Code Generation Algorithms

- Macro-expansion of IR operations (quadruples)
- "Simple code generation algorithm" (ALSU2e Section 8.6)
- Trade-off:
  - Registers vs. memory locations for temporaries
  - Sequencing
  - Code generation for expression trees
    - Labeling algorithm [Ershov 1958] [Sethi, Ullman 1970] (see later)

- Code generation using pattern matching
  - For trees: Aho, Johnson 1976 (dynamic programming), Graham/Glanville 1978 (LR parsing), Fraser/Hansson/Proebsting 1992 (IBURG tool), ...
  - For DAGs: [Ertl 1999], [K., Bednarski 2006] (DP, ILP)

Macro expansion of quadruples

- Each quadruple is translated to a sequence of one or several target instructions that performs the same operation.
  - very simple
  - bad code quality
    - Cannot utilize powerful instructions/addressing modes that do the job of several quadruples in one step
    - Poor use of registers

\( \rightarrow \) Simple code generation algorithm, see TDDB44/TDDD16 ([ALSU2e] 8.6)

Machine model (here: a simple register machine)

- **Register set**
  - E.g. 32 general-purpose registers R0, R1, R2, ...
  - some of them reserved (sp, fp, pc, sr, retval, par1, par2 ...)

- **Instruction set** with different addressing modes
  - Cost (usually, time / latency) depends on the operation and the addressing mode

Example: PDP-11 (CISC), instruction format \( \text{OP src, dest} \)

<table>
<thead>
<tr>
<th>Source operand</th>
<th>Destination address</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>register</td>
<td>register</td>
<td>1</td>
</tr>
<tr>
<td>register</td>
<td>memory</td>
<td>2</td>
</tr>
<tr>
<td>memory</td>
<td>register</td>
<td>2</td>
</tr>
<tr>
<td>memory</td>
<td>memory</td>
<td>3</td>
</tr>
</tbody>
</table>

Towards code generation by pattern matching

Example: Data flow graph (expression tree) for \( i = c + 4 \)

- in LCC-IR (DAGs of quadruples) [Fraser/Hanson ‘95]
- \( i, c \): local variables

In quadruple form:

- (Convention: last letter of opcode gives result type: \( I = \text{int}, C = \text{char}, P = \text{pointer} \))
- \( \text{ADDRLP, } I, 0, 11 \) \( \rightarrow 11 \rightarrow b+p+4 \)
- \( \text{ADDRLP, } C, 0, 12 \) \( \rightarrow 12 \rightarrow b+p+12 \)
- \( \text{CMSTI, 4, 0, 15} \) \( \rightarrow 13 \rightarrow M[2] \)
- \( \text{CMSTI, 4, 0, 15} \) \( \rightarrow 13 \rightarrow M[2] \)
- \( \text{ASGNI, } 8, 0, 11 \) \( \rightarrow 10 \rightarrow M[11] \)
Recall: Macro Expansion

- For the example tree:
  - s1, s2, s3...: "symbolic" registers (allocated but not assigned yet)
  - Target processor has delayed load (1 delay slot)

Using tree pattern matching...

- Utilizing the available addressing modes of the target processor, 3 instructions and only 2 registers are sufficient to cover the entire tree:

Tree patterns vs. Complex patterns

- Complex patterns
  - Forest patterns (several pattern roots)
  - DAG patterns (common subexpressions in pattern)

- Tree pattern (Multiply-add)
- Forest pattern (SIMD instruction)
- DAG pattern (Memory incr)

No match of mad (pending use of MUL)

Code generation by pattern matching

- Powerful target instructions/addressing modes may cover the effect of several quadruples in one step.
- For each instruction and addressing mode, define a pattern that describes its behavior in terms of quadruples resp. data-flow graph nodes and edges (usually limited to tree fragment shapes: tree pattern).
- A pattern matches at a node v if pattern nodes, pattern operators and pattern edges coincide with a tree fragment rooted at v
- Each instruction (tree pattern) is associated with a cost, e.g., its time behavior or space requirements
- Optimization problem: Cover the entire data flow graph (expression tree) with matching tree patterns such that each node is covered exactly once, and the accumulated cost of all covering patterns is minimal.

Tree grammar (Machine grammar)

The target processor is described by a tree grammar \( G = (N, T, \iota, P) \)

- \( N \) = set of nonterminals
- \( T \) = set of terminals
- \( \iota \) = start symbol (start symbol is start)
- \( P \) = list of production rules

Tree Grammar / Machine Grammar

Formally, we specify for each target machine a tree grammar \( G = (N, T, \iota, P) \)

- \( N \) = set of nonterminals
- \( T \) = set of terminals
- \( \iota \) = start symbol
- \( P \) = list of production rules
Derivation of the expression tree

Here: Top-down derivation

Derivation using a LR parser
(bottom-up derivation)

Some methods for tree pattern matching

- Use a LR-parser for matching [Graham, Glanville 1978]
  - compact specification of the target machine
  - using a context-free grammar ("machine grammar")
  - quick matching
  - not total-cost aware
  - greedy locally choices at reduce decisions → suboptimal
- Combine tree pattern matching with dynamic programming for total cost minimization
  [Aho, Ganapathi, Tjiang '89] [Fraser, Hanson, Proebsting '92]
- A LR parser is stronger than what is really necessary for matching tree patterns in a tree.
  - Right machine model is a tree automaton
  - a finite automaton operating on input trees rather than flat strings [Ferdinand, Seidl, Wilhelm '92]
- By Integer Linear Programming [Wilson et al. '94] [K., Bednarski '06]

Example: IBURG

Phase 1: bottom-up labeller
- annotates each node v of the input tree with
  - the set of tree patterns that match v and
  - their accumulated costs;
  - if multiple rules apply, pick one with locally minimum cost for each lhs nonterminal;
- Apply chain rules nonterm1 → nonterm2 as far as possible.
Example: IBURG

Phase 2: Top-down reducer
- Root of the labeled tree must correspond to start symbol (stmt)
- Choose best production for root node (accumulated costs), apply the corresponding productions,
- And do this recursively for each nonterminal in the rhs term

Example: IBURG

Found least-cost derivation:
Example: IBURG

Phase 3: Emitter
- in reverse order of the derivation found in phase 2:
  - emit the assembler code for each production applied
  - execute additional compiler code associated with these rules
    - e.g. register allocation.

Example: IBURG

Emitter result:

Example: IBURG

Given: a tree grammar describing the target processor
1. parse the tree grammar
2. generate:
   - bottom-up labeller,
   - top-down reducer,
   - emitter automaton
\rightarrow retargetable code generation!

Complexity of Tree Pattern Matching
- NP-complete if associativity / commutativity included
- Naive: \(O(\# \text{ tree patterns} \times \text{size of input tree})\)
- Preprocessing initial tree patterns
  - [Kron'75] [Hoffmann/O'Donnell'82]
    - may require exponential space / time
    - but then tree pattern matching in time \(O(\text{size of input tree})\)
- Theory of (non)deterministic tree automata
  - [Ferdinand/Seidl/Wilhelm'92]

Instruction selection for DAGs
Computing a minimal cost covering (with tree patterns) for DAGs?
- NP-complete [Proebsting'98]
  - For common subexpressions, only one of possibly several possible coverings can be realized.
  - Dynamic programming algorithm for trees OK as heuristic for regular processor architectures
  - The algorithm for trees may create optimal results for DAGs for special tree grammars (usually for regular register sets).
    - This can be tested a priori! [Ertl POPL'99]

Complex Patterns (1)
- Several roots possible
- Common subexpressions possible
  - SIMD instructions
  - DIVU instruction on Motorola 68K (simultaneous div + mod)
  - Read/Modify/Write instructions on IA32
  - Autocrement / autodecrement memory access instructions
- Min-cost covering of a DAG with complex patterns?
  - Can be formulated as PBQP instance [Scholz,Eckstein '03] (partitioned boolean quadratic programming)
  - Or as ILP (integer linear programming) instance
- Caution: Risk of creating artificial dependence cycles!
Complex Patterns (2)

**Caution:** Risk of creating artificial dependence cycles!

Example [Ebner 2009]:

```
*p := r+4;
*q := p+4;
*r  := q+4;
```

use postdecr. store instruct.:

```
ASGN
ADD
CNST #4
q
```

```
q' := q+4
```

```
ASGN
ADD
CNST #4
r
```

```
r' := r+4
```

Cycle between resulting instructions → No longer schedulable!

Solution [Ebner 2009]:

Add constraints to guarantee schedulability (some topological order exists)

---

Interferences with instruction scheduling and register allocation

- The cost attribute of a production is only a rough estimate
  - E.g., best-case latency or occupation time
- The actual impact on execution time is only known for a given scheduling situation:
  - currently free functional units
  - other instructions that may be executed simultaneously
  - latency constraints due to previously scheduled instructions
- Integration with instruction scheduling would be great!!

- *Mutations* with different unit usage may be considered:
  - a = 2*b equivalent to a = b<<1 and a = b+b (integer)
- Different instruction selections may result in different register need.

---

Retargetable Compilers

**Variant 1:** Use a Code Generator Generator

**Variant 2:** Parameterizable Code Generator

---

Some Literature on Instruction Selection