CSC 4181
Compiler Construction

Code Generation & Optimization

Code Generation

• Parser is modified to:
  - Output assembly language (or other low level) code for each statement
  - Use machine-specific knowledge
  - Perform simple or complex optimizations
  - Map statements to assembly code

Compiler Code Optimizations

• A code optimizer sits between the front end and the code generator.
  - Works with intermediate code.
  - Can do control flow analysis.
  - Can do data flow analysis.
  - Does transformations to improve the intermediate code.

Compiler Code Optimizations

• Optimizations provided by a compiler includes:
  - Inlining small functions
  - Code hoisting
  - Dead store elimination
  - Eliminating common sub-expressions
  - Loop unrolling
  - Loop optimizations: Code motion, Induction variable elimination, and Reduction in strength.

Another Way to Organize Them

• Local Optimizations
  - Constant Folding
  - Constant Propagation
  - Copy Propagation
  - Reduction in Strength
  - Inlining
  - Common sub-expression elimination
• Loop Optimizations
  - Loop Invariants
  - Reduction in strength due to induction variables
  - Loop unrolling
• Global Optimizations
  - Dead Code elimination
  - Code motion

Unnecessary Operations

• Common subexpression elimination
  - Save partial values instead of recomputing them
  - E.g. \( \frac{a^2}{a^2 + b^2} \)
• Dead code elimination
  - Don’t generate unreachable code
• Jump optimization
  - Avoid jumping to jump statements
• Useless code
  - Store X ; Load X
Costly Operations:
Constant Folding

• Don’t generate operations on constants; do them at compile time
  – Circ = 2 * PI * r
• If a variable will act like a constant for part or all of a program, transform its expressions also (constant propagation)

Constant Folding

• Subexpressions whose operands are all constants can be carried out at compile-time.
• E.g.
  X := 2 * 4
  Rather than generating
  Mov 2 r5
  Mov 4 r7
  Prim *** [5,7] r9
  Generate this instead
  Mov 8 r9

• Code like this is not ordinarily written by programmers but is often the result of translation of index calculations.

Constant Propagation

• Sometimes we know a symbolic value is a constant. So we can propagate the constant and generate better code:
  1 step := 4
  2 total := 0
  3 i := 1
  4 total := x [i + step]

• Note that because there are no possible paths to 4 that do not pass through 1, 2 and 3 we know that i+step can be computed by (1+4) which is known at compile time to be 5.

Copy Propagation

• Assignments of one variable to another also propagate information:
  x := y
  . . .
  total := Z[x]
• Note if my translation knows that y is stored in some register, R7, I can use R7 rather than fetching x from memory
• Copy propagation my remove all references to x completely. This allow makes the assignment to x dead code and a candidate to further optimization.

Costly Operations : Reduction in Strength

• Don’t call a power function if you can multiply
  – (x*x vs x^2)
• Don’t multiply when you can shift
  – (x<<4 vs. 16*x)

Reduction in Strength

• Some sequences of code can be replaced with simpler (or less expensive) sequences.
  x := 2 * y
  could be replaced by
  x := y + y
• Exponentiation by 2 by multiply
  – x ^2 == x * x
• Multiplication by factor of 2 by shift
Costly Operations: Function Calls

- Function calls have overhead
  - Create a new activation frame
  - Set up parameters & return value
  - Remove activation frame & jump to return
- Avoid by inlining (insert function code directly)
- Eliminate tail recursion with loop

Compiler Code Optimizations

- Inlining small functions
  - Repeatedly inserting the function code instead of calling it, saves the calling overhead and enable further optimizations.
  - Inlining large functions will make the executable too large.

Inlining

- Some calls to functions (especially primitives like +, -, *, absolute value, ord and char) can be inlined as a sequence of machine instructions instead of a call to a library routine.

```assembly
i := abs(j)
Bneg j L2
Mov j i
Br L3
L2:
Neg j R2
Mov R2 i
L3:
```

Common Sub-expressions

- Common subexpressions can be exploited by not duplicating code
  
x := z[j+2] - w[j+2]

  T1 := j+2
  x := z[T1] - w[T1]

Common Sub-expressions

- Note that common subexpressions often occur even when they are not in the user level code.
  - E.g. Subscript computations on two multi-dimensional arrays with the same dimensions will often have common sub expressions even if the index to the arrays are completely different

Loop Invariants

- Computations inside loops which remain invariant each time around the loop can be computed once outside the loop rather than each time around the loop.
- Note that index calculation may also introduces computations which are invariant around the loop.
Loop Invariants

For \( i := 1 \) to \( N \) do
\[
\{ \text{total} := \frac{x[i]}{\text{sqr}(n)} + \text{total} \}
\]

\( T_1 := \text{sqr}(n) \)

For \( i := 1 \) to \( N \) do
\[
\{ \text{total} := \frac{x[i]}{T_1} + \text{total} \}
\]

Induction Variables and reduction in strength

• Variables which vary in a regular way around loops are called induction variables.

• For loop variables are obvious cases, but implicit induction variables give much opportunity for optimization.

For \( i := 1 \) to \( 10 \) do
\[
\{ k := i \times 4; \text{total} := x[i] + w[k] \}
\]

Induction Variables (2)

• Note that \( k \) varies as a linear function of \( i \).

\[
\begin{align*}
i & := 1 \\
k & := 4 \\
\text{while } i & \leq 10 \text{ do} \\
& \{ \text{total} := x[i] + w[k] \\
& \quad i := i + 1; \quad k := k + 4 \}
\end{align*}
\]

Loop Unrolling

• Loop with low trip count can be unrolled. This does away with the loop initialization and test for termination conditions.

\[
\begin{align*}
\text{list} & := [1,2] \\
\text{while } \text{list} & \neq \text{nil} \text{ do} \\
& \{ \text{total} := \text{total} + \text{hd}(\text{list}) \\
& \quad \text{list} := \text{tail}(\text{list}) \}
\end{align*}
\]

Dead Code Elimination

• Automatic generation techniques often generate code that is unreachable.

\[
\begin{align*}
\text{debug} & := \text{false}; \\
\text{if } \text{debug} & \quad \text{then print } x; \quad f(x);
\end{align*}
\]

• Because of constant propagation it is possible to tell at compile-time that the then branch will never be executed.

Compiler Code Optimizations

• Dead store elimination
  – If the compiler detects variables that are never used, it may safely ignore many of the operations that compute their values.
Code hoisting

- Code hoisting
  - Moving computations outside loops
  - Saves computing time

Code Hoisting (2)

- Code hoisting
  - In the following example (2.0 * PI) is an invariant expression there is no reason to recompute it 100 times.
    
    DO I = 1, 100
    ARRAY(I) = 2.0 * PI * I
    ENDDO
  
  - By introducing a temporary variable 't' it can be transformed to:
    
    t = 2.0 * PI
    DO I = 1, 100
    ARRAY(I) = t * I
    ENDDO

Code Motion (reordering)

- Sometimes reordering statements that do not interfere, allows other more powerful optimizations to be come applicable.

    Push R2
    Mov I R3
    Pop R4

    Mov I R3
    Push R2
    Pop R4

    Mov I R3
    Mov R2 R4

- Now copy propagation might remove R2 altogether

Code Motion (Code Hoisting)

- Branches in code sometimes repeat identical calculations.
- These calculations can sometimes be “hoisted” before the branch, then they don’t have to be repeated.
- This saves space, but not time.

    if g(x)
      then x := (d*2) + w / k
      else x := (d*2) - w / j
    T1 := (d*2);
    if g(x)
      then x := T + w / k
      else x := T - w / j

- Multi branch “case” statements can make this quite a space saver

Register Allocation

- Task: Manage scarce resources (registers) in environment with imperfect information (static program text) about dynamic program behavior.
- General aim is to keep frequently-used values in registers as much as possible, to lower memory traffic. Can have a large effect on program performance.
- Variety of approaches are possible, differing in sophistication and in scope of analysis used.

Spilling

- Allocator may be unable to keep every “live” variable in registers; must then “spill” variables to memory. Spilling adds new instructions, which often affects the allocation analysis, requiring a new iteration.
- If spilling is necessary, what should we spill? Some heuristics:
  - Don’t spill variables used in inner loops.
  - Spill variables not used again for “longest” time.
  - Spill variables which haven’t been updated since last read from memory.
Simplistic approach
• Assume variables "normally" live in memory.
• Use existing (often redundant) fetches and stores present in IR1.
• So: only need to allocate registers to IR temporaries (T5 etc.).
• Ignore possibility of spills.
• Use simple linear scan register allocator based on liveness intervals.

Liveness
• To determine how long to keep a given variable (or temporary) in a register, need to know the range of instructions for which the variable is live.
• A variable or temporary is live immediately following an instruction if its current value will be needed in the future (i.e., it will be used again, and it won't be changed before that use).

Example

```
| instruction:          | live after |
T2 := 3                | T2         |
T3 := T2               | T2         |
T4 := T3 + 4           | T2 T3      |
T4 := T2 + T4          | T2         |
a := T4                | (nothing)  |
```