CS 526
Advanced Compiler Construction

https://charithm.web.illinois.edu/cs526/sp2022/
(slides adapted from Sasa and Vikram)
Brief Announcements

• Project 1 is out.
  • Deadline 2/12 by 11.59pm CST
  • Tip: Start soon since it is a coding assignment.
• Miniquiz grading policy
  • 20 quizzes. Can skip 5 without penalty.
  • Extra credit for each one done extra.
  • ~0.67% per quiz broken down into:
    – 50% - correct solution
    – 50% - participation
STATIC SINGLE ASSIGNMENT

The slides adapted from Vikram Adve
SSA Construction Algorithm

Steps:
1. Compute the dominance frontiers*
2. Insert \( \varphi \)-functions
3. Rename the variables

Thm. Any program can be put into minimal SSA form using the previous algorithm. [Refer to the paper for proof]
Control Dependence

Def. Postdomination: node \( p \) postdominates a node \( d \) iff all paths to the exit node of the graph starting at \( d \) must go through \( p \).

Def. In a CFG, node \( Y \) is control-dependent on node \( B \) if
- There is a path \( B,N_1,N_2, ..., N_k,Y \), such that \( Y \) postdominates \( N_1 \ldots N_k \) (possibly empty), and
- \( Y \) does not postdominate \( B \).

Def. Reverse Control Flow Graph (RCFG) of a CFG has the same nodes as CFG and has edge \( Y \rightarrow X \) if \( X \rightarrow Y \) is an edge in CFG.
- \( p \) is a postdominator of \( d \) iff \( p \) dominates \( d \) in the RCFG.
SSA-Based Optimizations

- Dead Code Elimination (DCE)
- Sparse Conditional Constant Propagation (SCCP)
- Loop-Invariant Code Motion (LICM)
- Global Value Numbering (GVN)
- Strength Reduction of Induction Variables
- Live Range Identification in Register Allocation
Constant Propagation

Goals
Whenever there is a statement of the form \( v = \text{Const} \), the uses of \( v \) can be replaced by \( \text{Const} \).

Safety
Analysis: Explicit propagation of constant expressions
Transformation: Most languages allow removal of computations

Profitability
Fewer computations, almost always

Opportunity
Symbolic constants, conditionally compiled code, …
Simple Constant Propagation

Worklist = All statements in the SSA program

While Worklist ≠ ∅
    remove a statement S from Worklist

    if S is “v = φ(c1,...cn)” and c1=...=cn=c (Const),
        replace S with v = c

    if S is “v = c” (c is Const)
        Delete S from the program
    For each Statement T ∈ Uses (v)
        substitute v with c in T
    Worklist = Worklist ∪ {T}
Extensions of the Algorithm

Copy propagation:

• Assignments $x = y$ or $x = \varphi(y)$ can be replaced by a simple use of $y$.

Constant folding:

• Assignments of the form $x = a \odot b$ can be immediately evaluated if $a$ and $b$ are constants, and the statement replaced with $x = c$ ($c = a \odot b$)

Constant conditions:

• If a condition if $(x \odot y)$ always evaluate to true or false, then keep only one branch.
Conditional Constant Propagation: SCCP

Goals
Identify and replace SSA variables with constant values
Delete infeasible branches due to discovered constants

Safety
Analysis: Explicit propagation of constant expressions
Transformation: Most languages allow removal of computations

Profitability
Fewer computations, almost always (except pathological cases)

Opportunity
Symbolic constants, conditionally compiled code, ...
Example 1

J = 1;
...
if (J > 0)
    I = 1; // Always produces 1
else
    I = 2;
Example 2

I = 1;
...
while (...) {
    J = I;
    I = f(...);
...
    I = J; // Always produces 1
}

We need to proceed with the assumption that everything is constant until proved otherwise.
Example 3

I = 1;
...
while (...) {
    J = I;
    I = f(...);
    ...
    if (J > 0)
        I = J; // Always produces 1
}
Conditional Constant Propagation*

**Advantage:**
Simultaneously finds constants + eliminates infeasible branches.

**Optimistic**
Assume every variable may be constant, until proven otherwise. (Pessimistic $\equiv$ initially assume nothing is constant.)

**Sparse:** Only propagates variable values where they are actually used or defined (using def-use chains in SSA form).

**Iterative:**
Build the list of constant definitions and uses using a worklist algorithm.

---

*sccep: Sparse Conditional Constant Propagation

Sparse conditional constant propagation and merging, which can be summarized as:

- Assumes values are constant unless proven otherwise
- Assumes BasicBlocks are dead unless proven otherwise
- Proves values to be constant, and replaces them with constants
- Proves conditional branches to be unconditional

Note that this pass has a habit of making definitions be dead. It is a good idea to run a DCE pass sometime after running this pass.

* Constant Propagation with Conditional Branches; M. Wegman and K. Zadeck, TOPLAS’91
Dead Code Elimination

The results of the computation are visible through return values or output statements

- We can remove the instructions that do not contribute to the visible outputs

A simple algorithm:

- Compute (or maintain) the def-use chains
- Iterate over the instructions:
  - For $v = x \text{ op } z$, if $v$ has at least one use, mark as live, otherwise mark as dead
- Remove the instructions marked as dead
Aggressive Dead Code Elimination

Ordinary DCE:
- \( x_2 \) is live because used in def. of \( x_1 \).
- \( x_1 \) is live because used in def. of \( x_2 \).
Yet…

Idea: Analogous to SCCP, be optimistic and assume a statement is dead unless proven otherwise (i.e., contributes to the output)

Transform Passes

This section describes the LLVM Transform Passes.

**-adce**: Aggressive Dead Code Elimination

ADCE aggressively tries to eliminate code. This pass is similar to **DCE** but it assumes that values are dead until proven otherwise. This is similar to **SCCP**, except applied to the liveness of values.
Aggressive Dead Code Elimination

In each step, mark a statement as live if:

1. It is the **output** statement (e.g., return)
2. It has **known side effects** (e.g., assignment to global variable or calling a function with side effects)
3. It **defines a variable** $x$ used by an already live statement
4. It is a **conditional branch**, and some other, already live statement is **control dependent** on the branch (and its block)

The algorithm then converges to a set of live variables

- Caveat: the algorithm may remove “empty” infinite loops
Loop-Invariant Code Motion

Example:

```c
x = 1; y = 0
while ( y < 10 ) {
    t = min(x,2)
    y = y + x
}
```

Becomes:

```c
x = 1; y = 0
t = min(x,2)
while ( y < 10 ) {
    y = y + x
}
```

Pattern:

```c
v = a op b
loop {
    ... v is used here
}
```

Becomes:

```c
v = a op b
loop {
    ... v is used here
}
```

What conditions does the code need to satisfy for this transformation to be sound?
Loop-Invariant Code Motion

**Analysis**

Conditions for the analysis ($v = a \text{ op } b$):

- **Both $a$, $b$ are constants**
  
  ```
  while (b) { v = 2 + 3; /* ... */ }
  ```

- **Both $a$, $b$ are defined before the loop (SSA ensures there is only a single dominating definition for each)**
  
  ```
  x = ...;
  while (b) { v = x + 1; /* ... */ }
  ```

- **Both $a$, $b$ are referring to the variables that are in the loop but already determined to be loop invariant**
  
  ```
  x = ...;
  while (b) { v = x + 1; t = v * 2; /* ... */ }
  ```

If curious about what complications arise if the program is not in the SSA form see [http://www.cs.cmu.edu/~aplatzer/course/Compilers11/17-loopinv.pdf](http://www.cs.cmu.edu/~aplatzer/course/Compilers11/17-loopinv.pdf)
Loop-Invariant Code Motion

Transformation

**Version 1:** Since the computation is in SSA form, just move it to the node before the header

- What if there is no single such node? (Make it!)
- **Loop preheader:** a single node that dominates the loop header
- How do we ensure there are no side effects?
- What if the loop-invariant computation is expensive?

**Version 2:** Like candidate 1, but add loop’s condition:

```c
if (cond) {
    t = a op b;
    while (cond) { /* ... */ }
}
```

---

**-licm:** Loop Invariant Code Motion

This pass performs loop invariant code motion, attempting to remove as much code from the body of a loop as possible. It does this by either hoisting code into the preheader block, or by sinking code to the exit blocks if it is safe. This pass also promotes mutually aliased memory locations in the loop to live in registers, thus hoisting and sinking “invariant” loads and stores.
Another Handy One...

-loop-simplify: Canonicalize natural loops

This pass performs several transformations to transform natural loops into a simpler form, which makes subsequent analyses and transformations simpler and more effective. A summary of it can be found in Loop Terminology, Loop Simplify Form.

Loop pre-header insertion guarantees that there is a single, non-critical entry edge from outside of the loop to the loop header. This simplifies a number of analyses and transformations, such as LICM.

Loop exit-block insertion guarantees that all exit blocks from the loop (blocks which are outside of the loop that have predecessors inside of the loop) only have predecessors from inside of the loop (and are thus dominated by the loop header). This simplifies transformations such as store-sinking that are built into LICM.

This pass also guarantees that loops will have exactly one backedge.

Note that the simplifycfg pass will clean up blocks which are split out but end up being unnecessary, so usage of this pass should not pessimize generated code.

This pass obviously modifies the CFG, but updates loop information and dominator information.

See also: https://llvm.org/docs/LoopTerminology.html
Induction Variable Substitution

Auxiliary Induction Variable
An auxiliary induction variable in a loop

```c
for (int i = 0; i < n; i++) {
    ... 
}
```
is any variable \( j \) that can be expressed as

\[
c \times i + m
\]
at every point where it is used in the loop, where \( c \) and \( m \) are loop-invariant values, but \( m \) may be different at each use.
Optimization Goals

Identify linear expression for each auxiliary induction variable

• More effective dependence analysis, loop transformations
• Substitute linear expression in place of every use
• Eliminate expensive or loop-invariant operations from loop
Induction Variable Substitution

Auxiliary Induction Variable

```c
for (int i = 0; i < n; i++) {
    j = 2*i + 1;
    k = -i;
    l = 2*i*i + 1;
    c = c + 5;
}
```
Induction Variable Substitution

Auxiliary Induction Variable

for (int i = 0; i < n; i++) {
    j = 2*i + 1;        // Y
    k = -i;            // Y
    l = 2*i*i + 1;     // N
    c = c + 5;         // Y*
}
Reminder: Strength Reduction

**Goal:** Replace expensive operations by cheaper ones

**Primitive Operations:** Many Examples

\[
\begin{align*}
n \times 2 & \rightarrow n \ll 1 \text{ (similarly, } n/2) \\
n \times 2 & \rightarrow n \times n
\end{align*}
\]

**Recurrences**

Example: \( x = a[i] \) to \( x = (\text{base}(a) + (i-1) \times 4) \)

Such recurrences are common in array address calculations
Induction Variable Substitution

Strategy

• Identify operations of the form:
  \[ x \leftarrow \text{iv} \times c \quad \text{or} \quad x \leftarrow \text{iv} \pm c \]
  iv: induction variable or another recurrence
  c: loop-invariant variable

• Eliminate **multiplications** from the loop body

• Eliminate induction variable if the **only remaining use** is in the loop **termination test**

-indvars: Canonicalize Induction Variables

This transformation analyzes and transforms the induction variables (and computations derived from them) into simpler forms suitable for subsequent analysis and transformation.
Induction Variable Substitution

do i = 1 to 100
    sum = sum + a(i)
enddo

Source code

sum = 0.0
i = 1
L:
    t1 = i - 1
    t2 = t1 * 4
    t3 = t2 + a
    t4 = load t3
    sum = sum + t4
    i = i + 1
    if (i <= 100) goto L

Intermediate code

sum0 = 0.0
i0 = 1
L:
    sum1 = \phi(sum0, sum2)
    i1 = \phi(i0, i2)
    t10 = i1 - 1
    t20 = t10 * 4
    t30 = t20 + a
    t40 = load t30
    sum2 = sum1 + t40
    i2 = i1 + 1
    if (i2 <= 100) goto L

SSA form
Induction Variable Substitution

\[
\begin{align*}
\text{sum}_0 &= 0.0 \\
\text{i}_0 &= 1 \\
\text{L:} & \quad \text{sum}_1 = \phi(\text{sum}_0, \text{sum}_2) \\
& \quad \text{i}_1 = \phi(\text{i}_0, \text{i}_2) \\
& \quad \text{t}_10 = \text{i}_1 - 1 \\
& \quad \text{t}_20 = \text{t}_10 \times 4 \\
& \quad \text{t}_30 = \text{t}_20 + a \\
& \quad \text{t}_40 = \text{load } \text{t}_30 \\
& \quad \text{sum}_2 = \text{sum}_1 + \text{t}_40 \\
& \quad \text{i}_2 = \text{i}_1 + 1 \\
& \quad \text{if } (\text{i}_2 \leq 100) \text{ goto } \text{L} \\
\end{align*}
\]

SSA form

\[
\begin{align*}
\text{sum}_0 &= 0.0 \\
\text{i}_0 &= 1 \\
\text{L:} & \quad \text{t}_50 = a \\
& \quad \text{sum}_1 = \phi(\text{sum}_0, \text{sum}_2) \\
& \quad \text{i}_1 = \phi(\text{i}_0, \text{i}_2) \\
& \quad \text{t}_51 = \phi(\text{t}_50, \text{t}_52) \\
& \quad \text{t}_40 = \text{load } \text{t}_50 \\
& \quad \text{sum}_2 = \text{sum}_1 + \text{t}_40 \\
& \quad \text{i}_2 = \text{i}_1 + 1 \\
& \quad \text{t}_52 = \text{t}_51 + 4 \\
& \quad \text{if } (\text{i}_2 \leq 100) \text{ goto } \text{L} \\
\end{align*}
\]

After strength reduction
Induction Variable Substitution

\[
\begin{align*}
\text{sum}_0 &= 0.0 \\
i_0 &= 1 \\
t_{50} &= a \\
L: & \text{ sum}_1 = \phi(\text{sum}_0, \text{sum}_2) \\
i_1 &= \phi(i_0, i_2) \\
t_{51} &= \phi(t_{50}, t_{52}) \\
t_{40} &= \text{load } t_{50} \\
\text{sum}_2 &= \text{sum}_1 + t_{40} \\
i_2 &= i_1 + 1 \\
t_{52} &= t_{51} + 4 \\
\text{if } (i_2 \leq 100) \text{ goto L}
\end{align*}
\]

After strength reduction

\[
\begin{align*}
\text{sum}_0 &= 0.0 \\
t_{50} &= a \\
L: & \text{ sum}_1 = \phi(\text{sum}_0, \text{sum}_2) \\
t_{51} &= \phi(t_{50}, t_{52}) \\
t_{40} &= \text{load } t_{50} \\
\text{sum}_2 &= \text{sum}_1 + t_{40} \\
t_{52} &= t_{51} + 4 \\
\text{if } (t_{52} \leq 396 + a) \text{ goto L}
\end{align*}
\]

After induction variable substitution
Induction Variable Substitution (recap)

\[
\begin{align*}
    \text{sum}_0 &= 0.0 \\
    t_{50} &= a \\
    L:\quad \text{sum}_1 &= \phi(\text{sum}_0,\text{sum}_2) \\
    t_{51} &= \phi(t_{50},t_{52}) \\
    t_{40} &= \text{load } t_{50} \\
    \text{sum}_2 &= \text{sum}_1 + t_{40} \\
    t_{52} &= t_{51} + 4 \\
    \text{if } (t_{52} \leq 396 + a) \text{ goto } L
\end{align*}
\]

After induction variable substitution

\[
\begin{align*}
    \text{sum}_0 &= 0.0 \\
    i_0 &= 1 \\
    L:\quad \text{sum}_1 &= \phi(\text{sum}_0,\text{sum}_2) \\
    i_1 &= \phi(i_0,i_2) \\
    t_{10} &= i_1 - 1 \\
    t_{20} &= t_{10} \times 4 \\
    t_{30} &= t_{20} + a \\
    t_{40} &= \text{load } t_{30} \\
    \text{sum}_2 &= \text{sum}_1 + t_{40} \\
    i_2 &= i_1 + 1 \\
    \text{if } (i_2 \leq 100) \text{ goto } L
\end{align*}
\]

Before induction variable substitution
References

Cocke and Kennedy, CACM 1977 (superseded by the next one).
Allen, Cocke and Kennedy, “Reduction of Operator Strength,” In

Classical Approach

• ACK: Classic algorithm, widely used.
• works on “loops” (Strongly Connected Regions) of flow graph
• uses def-use chains to find induction variables and recurrences


SSA-based algorithm

• Same effectiveness as ACK, but faster and simpler
• Identify induction variables from SCCs in the SSA graph
Value Numbering

Code:
\[
\begin{align*}
a &= x + y \\
b &= x + y \\
a &= 1 \\
c &= x + y \\
d &= y + x \\
e &= d - 1 \\
f &= e + 1
\end{align*}
\]

- **Analysis**: Determining equivalent computations (variables, expressions, consts)
- **Transformation**: Eliminates duplicates with a semantics-preserving optimization
- Form of redundancy elimination
Value Numbering

- Assign an **identifying number** to each variable / expression / constant:
  
  \[ x \text{ and } y \text{ have same id number } \iff x = y \text{ for all inputs} \]

- Use algebraic identities to simplify expressions
- Discover redundant computations & replace them
- Discover constant values, fold & propagate them
Value Numbering

• Use algebraic identities to simplify expressions
  • Commutativity \((a+b = b+a), (a+b+c = c+b+a), (a+b)^2 = a^2+2ab+b^2\ldots\)

• Discover redundant computations and replace them
  • E.g., \(y=2\times x;\ z=2\times x+1 \Rightarrow y=2\times x;\ z=y+1\)

• Discover constant values, fold & propagate them
  • After SCCP: e.g., \(x=1;\ y = x+1 \Rightarrow y = 1+1\)
  • Evaluate constant expression \((y = 2)\) then propagate
Local Value Numbering

- Each variable, expression, & constant gets a “value number” (hash code)

  **Same value number ⇒ same value**

- **Prerequisites:** low-level intermediate code and existing basic blocks
- Equivalence based solely on facts from within the single basic block
- If an instruction’s value number is already defined, instr. can be eliminated & subsequent references subsumed
- Constant folding is simple
Local Value Numbering

\[ a = x + y \]
\[ b = x + y \]
\[ a = 1 \]
\[ c = x + y \]
\[ d = y + x \]
\[ e = d - 1 \]
\[ f = e + 1 \]

\[ V1 \leftarrow \text{hash}(+, VN[x], VN[y]), \]
\[ \text{Name}[V1] \leftarrow a \]
\[ \text{hash}(+, VN[x], VN[y]) == V1 \]
So, replace \( x+y \) with \( a \). Transformed: \( b = a \)

\[ \text{Name}[V1] \leftarrow \emptyset \quad \text{(can we be more precise?)} \]

Can we replace?

Challenges:
tracking where each value resides
commutativity \( \Rightarrow \) ???
identities (e.g., \( Vx \) OR \( Vx \times 1 \)): \( \Rightarrow \)
\[ \text{instr. gets value number of operand (Vx)} \]
Local Value Numbering

\[ a_1 = x + y \]
\[ b = x + y \]
\[ a_2 = 1 \]
\[ c = x + y \]
\[ d = y + x \]
\[ e = d - 1 \]
\[ f = e + 1 \]

\[ V_1 \leftarrow \text{hash}(+, \text{VN}[x], \text{VN}[y]), \]
\[ \text{Name}[V_1] \leftarrow a \]
\[ \text{hash}(+, \text{VN}[x], \text{VN}[y]) == V_1 \]
So, replace \( x+y \) with \( a \). Transformed: \( b = a_1 \)
\[ \text{Name}[V_1] \leftarrow \emptyset \quad \text{(don’t need anymore)} \]

Challenges:
tracking where each value resides
commutativity \( \Rightarrow \) ???
identities (e.g., \( V_x \text{ OR } V_x \times 1 \)):\( \Rightarrow \)
\[ \text{instr. gets value number of operand} \ (V_x) \]
Local Value Numbering

\[ a_1 = x + y \]
\[ b_1 = a_1 \]
\[ a_2 = 1 \]
\[ c_1 = a_1 \]
\[ d_1 = a_1 \]
\[ e = d_1 - 1 \]
\[ f = e + 1 \]

\[ V_l \leftarrow \text{hash}(+, \text{VN}[x], \text{VN}[y]), \]
\[ \text{Name}[V_l] \leftarrow a \]
\[ \text{hash}(+, \text{VN}[x], \text{VN}[y]) = V_l \]

So, replace \( x+y \) with \( a \). Transformed: \( b = a_1 \)

\[ \text{Name}[V_l] \leftarrow \emptyset \quad \text{(don't need anymore)} \]

Challenges:

What happens with \( e \) and \( f \)?
Local Value Numbering

For each instruction $i : x \leftarrow y \text{ op } z$ in the block

$V1 \leftarrow VN[y]$
$V2 \leftarrow VN[z]$

let $v = \text{hash(op, V1, V2)}$
if ($v$ exists in hash table)
  replace RHS with $\text{Name}[v]$
else
  enter $v$ in hash table
  $VN[x] \leftarrow v$
  $\text{Name}[v] \leftarrow \text{ti}$ (new temporary)
replace instruction with: “$\text{ti} \leftarrow y \text{ op } z; x \leftarrow \text{ti}$”
Local VN  *Simplifications*

- If the operands have the same value number i.e. \( z = x \ op \ y \), and \( \text{VN}[x] = \text{VN}[y] \)
  - if \( op \) is MAX, MIN, AND, OR, . . . replace \( op \) with a copy operation \( (z=x) \)
  - if \( op \) tests equality, replace it with \( z = \text{true} \)
  - if \( op \) tests inequality replace it with \( z = \text{false} \)
- if all operands \( (x,y) \) are constants and we haven’t already simplified the expression, then immediately evaluate the resulting constant and propagate constants down
- if one operand is constant and we haven’t yet simplified the expression:
  - if a constant operand is zero, replace ADD and OR with another operand; replace MULT, AND with zero
  - if constant operand is one, replace MULT with assignment of another operand
- If \( op \) commutes, reorder its operands into **ascending order by value number** (canonical form)
Local VN Analogy

- Constructing a DAG from a forest (set of trees)
- Each expression is a node in a dag, edges are uses of the expression in the instructions
- Start from the leading instruction of the basic block
- Collapse nodes that are repeated into a single node and connect the edges to all uses

\[
\begin{align*}
a &= x + y \\
b &= (x + y) - z \\
c &= y + x
\end{align*}
\]
Global Value Numbering

\[ W = X + Y; \]
\[ \text{if (\ldots)} \{ \]
\[ \quad Z = X + Y; \]
\[ \quad X = 1; \]
\[ \} \text{ else } \{ \]
\[ \quad Z = X + Y - 1; \]
\[ \} \]

\[ U = X + Y - 1; \quad // \ ?? \]
Global Value Numbering

\[ W1 = X1 + Y1; \]
\[ \text{if} \ (\ldots) \{ \]
\[ \quad Z1 = X1 + Y1; \]
\[ \quad X2 = 1; \]
\[ \} \ \text{else} \{ \]
\[ \quad Z2 = X1 + Y1 - 1; \]
\[ \} \]
\[ X3 = \Phi(X1, X2) \]
\[ Z3 = \Phi(Z1, Z2) \]
\[ U1 = X3 + Y1 - 1; \quad \text{// ??} \]
Global Value Numbering

\[
T_1 = X_1 + Y_1; \quad W_1 = T_1;
\]
\[
\text{if (\ldots) } \{
    Z_1 = T_1;
    X_2 = 1;
\}
\text{else } \{
    Z_2 = T_1 - 1;
\}
\]
\[
X_3 = \Phi(X_1, X_2)
\]
\[
Z_3 = \Phi(Z_1, Z_2)
\]
\[
U_1 = X_3 + Y_1 - 1; \quad // \quad ??
\]
Global Value Numbering

W = X + Y;
if (...) {
    Z = X + Y;
    W = 1 + Z;
} else {
    Z = X + Y - 1;
}

U = X + Y - 1;   // ??
Global Value Numbering

\[
W1 = X1 + Y1;
\]

\[
\text{if} \ (\ldots) \ {\}
\]
\[
\quad Z1 = X + Y;
\]
\[
\quad W2 = 1 + Z1;
\]
\[
} \ \text{else} \ {\}
\]
\[
\quad Z2 = X + Y - 1;
\]
\[
} \]
\[
W3 = \Phi(W1, W2)
\]
\[
Z3 = \Phi(Z1, Z2)
\]
\[
U1 = X1 + Y1 - 1; \quad // \quad ??
\]
Global Value Numbering

```
T1 = X1 + Y1; W1 = T1;
if (...) {
    Z1 = X1 + Y1;
    W2 = 1+Z1;
} else {
    Z2 = X1 + Y1 - 1;
}
W3 = Phi(W1, W2)
Z3 = Phi(Z1, Z2)
U1 = X1 + Y1 - 1;  // ??
```
Global Value Numbering

\[ T_1 = X_1 + Y_1; \quad W_1 = T_1; \]
if (...) {
\[ Z_1 = T_1; \]
\[ W_2 = 1 + Z_1; \]
} else {
\[ Z_2 = T_1 - 1; \quad // \quad X_1 + Y_1 - 1 \]
}
\[ W_3 = \text{Phi}(W_1, W_2) \]
\[ Z_3 = \text{Phi}(Z_1, Z_2) \]
\[ U_1 = X_1 + Y_1 - 1; \quad // \quad ?? \]
Global Value Numbering

\[
T1 = X1 + Y1; \quad W1 = T1;
\]
\[
\text{if (\ldots) \{} \\
\quad Z1 = T1; \\
\quad W2 = 1+Z1; \\
\text{\} else \{} \\
\quad Z2 = T1 - 1; \quad // \quad X1 + Y1 - 1 \\
\text{\} }
\]
\[
W3 = Phi(W1, W2) \\
Z3 = Phi(Z1, Z2) \\
U1 = T1 - 1; \quad // \quad ?? (Done?)
\]
Yet another example

\[
X_0 = 1 \\
Y_0 = 1 \\
\text{while ( . . . ) } \{ \\
\quad X_1 = \phi(X_0, X_2) \\
\quad Y_1 = \phi(Y_0, Y_2) \\
\quad X_2 = X_1 + 1 \\
\quad Y_2 = Y_1 + 1 \\
\}
\]
Global Value Numbering (DVNT)

The Dominator-based VN Technique (DVNT)

- B2, B3 can be value-numbered using B1’s table
- How about B4? Yes, can use the expressions from B1 (dominator node) but needs to invalidate the expressions killed in B2, B3
- Still based on hashing

**BUT:** difficult to merge these tables
  - A variable may be redefined in B2, B3, or both
Instruction Congruence

Instructions $i$ and $j$ are **congruent** iff:

1. They are the same instruction, or

2. They are assignments of constants, which are equal (e.g. $x := c_i$, $y := c_j$ and $c_i == c_j$), or

3. They have one or multiple operands, e.g.,

   $z_i = x_i \ op \ y_i$
   
   $z_j = x_j \ op \ y_j$

   *same* operator and their operands are **congruent** ($x_i$ congruent to $x_j$ and $y_i$ congruent to $y_j$), taking into consideration commutativity of $op$. 
A Global Approach (Alpern, Wegman & Zadeck)

Prerequisite: Computation must be in SSA Form

Algorithm:
1. partition instructions into congruence classes by opcode
2. worklist ← all classes
3. while (worklist is not empty)
   a) remove a class c from worklist
   b) for each class s that uses some \( x \in c 
      i. split s into \( s^c \) & \( s^{-c} \): all users of c are in \( s^c \) and those that are not are in \( s^{-c} \)
      ii. if s was on the worklist, remove s, \( s^c \) and \( s^{-c} \)
      iii. Else put smaller of \( s^c \) or \( s^{-c} \) onto the worklist
4. pick a representative instruction for each partition and perform replacement
Example

Congruence Classes:

1. 
   \( (Y_0 = X_0 + 1, \ Z_0 = X_0 + 1) \)

2. 
   \( (Y_1 = X_0 + 2, \ Z_1 = X_0 + 2) \)

3. 
   \( (Y_2 = \varphi(Y_0, Y_1), \ Z_2 = \varphi(Z_0, Z_1)) \)
Example

\[ X_0 \leq 1 \]

\[ Y_0 = X_0 + 1 \]
\[ Z_0 = Y_0 \]

\[ Y_1 = X_0 + 2 \]
\[ Z_1 = X_0 + 2 \]

\[ Y_2 = \varphi(Y_0, Y_1) \]
\[ Z_2 = \varphi(Z_0, Z_1) \]
\[ X_0 \leq 3 \]

Congruence Classes:

1. \( Y_0 = X_0 + 1, \ Z_0 = X_0 + 1 \)
2. \( Y_1 = X_0 + 2, \ Z_1 = X_0 + 2 \)
3. \( Y_2 = \varphi(Y_0, Y_1), \ Z_2 = \varphi(Z_1, Z_2) \)
Example

Congruence Classes:

1. $(Y_0 = X_0 + 1, Z_0 = X_0 + 1)$
2. $(Y_1 = X_0 + 2, Z_1 = X_0 + 2)$
3. $(Y_2 = \phi(Y_0, Y_1), Z_2 = \phi(Z_0, Z_1))$
Example

\[
\begin{align*}
X_0 &< 1 \\
Y_0 &= X_0 + 1 \\
Z_0 &= Y_0 \\
Y_1 &= X_0 + 2 \\
Z_1 &= Y_1 \\
Y_2 &= \phi(Y_0, Y_1) \\
Z_2 &= Y_2 \\
&< 3
\end{align*}
\]

Congruence Classes:

1. \((Y_0 = X_0 + 1, Z_0 = X_0 + 1)\)
2. \((Y_1 = X_0 + 2, Z_1 = X_0 + 2)\)
3. \((Y_2 = \phi(Y_0, Y_1), Z_2 = \phi(Z_1, Z_2))\)
Properties of the Algorithm

• Cannot prove congruences that involve different operators:
  - \(5 \times 2 \sim = 7 + 3\) or
  - \(3 + 1 \sim = 2 + 2\) or
  - \(x \times 1 \sim = x\)

• Need separate pass to transform code (partitioning must complete first)

• Powerful technique but ignores compile-time costs

• Alternative: SCC Based Algorithm (see references)
  - SCC often beats AWZ in practice
References

Long history in literature
• form of redundancy elimination (compare CSE)
• local version using hashing: late 60’s Cocke & Schwartz, 1969
• algorithms for blocks, extended blocks, dominator regions, entire procedures, and (maybe) whole programs
• easy to understand algorithm for single block
• larger scopes cause more complex algorithms


-gvn: Global Value Numbering

This pass performs global value numbering to eliminate fully and partially redundant instructions. It also performs redundant load elimination.
Optimizations where we will need more information

- Copy Propagation
- Global Common Subexpression Elimination (GCSE)
- Partial Redundancy Elimination (PRE)
- Redundant Load Elimination
- Dead or Redundant Store Elimination
- Code Placement Optimizations