CSE110A: Compilers

Topics:

• Start of Module 4 - Optimizations

Module 4: Optimizations

• What are compiler optimizations?

• Why do we want compiler optimizations?

- What are compiler optimizations?
 - automated program transforms designed to make code more optimal
 - optimal can mean different things
 - code optimized for one system might be different for code optimized for a different system
 - we can optimize for speed, for energy efficiency, or for code size. What else?
- Why do we want the compiler to help us optimize?
 - So we can write more maintainable/portable code
 - So we don't have to worry about learning nuanced details about every possible system

• What are some compiler optimizations you know about?

• What are some compiler optimizations you know about?

```
for (int i = 0; i < 10; i++) {
   x = x + 1;
}</pre>
```

loop unrolling

```
for (int i = 0; i < 10; i++) {
    x = x + 1;
    i++;
    x = x + 1;
}</pre>
```

• What are some compiler optimizations you know about?

```
for (int i = 0; i < 10; i++) {
  x = x + 1;
}</pre>
```

loop unrolling

```
for (int i = 0; i < 10; i++) {
    x = x + 1;
    i++;
    x = x + 1;
}</pre>
```

```
int foo() {
  int i,j,k;
  i = 10;
  j = i;
  k = j;
  return k;
}
```

constant propagation

```
int foo() {
   int i,j,k;
   return 10;
}
```

• What are some compiler optimizations you know about?

```
for (int i = 0; i < 10; i++) {
   x = x + 1;
}</pre>
```

loop unrolling

```
for (int i = 0; i < 10; i++) {
    x = x + 1;
    i++;
    x = x + 1;
}</pre>
```

What does this save us?

• What are some compiler optimizations you know about?

```
for (int i = 0; i < 10; i++) {
   x = x + 1;
}</pre>
```

loop unrolling

```
for (int i = 0; i < 10; i++) {
    x = x + 1;
    i++;
    x = x + 1;
}</pre>
```

optimizations at one stage can enable optimizations at another stage:

```
for (int i = 0; i < 10; i+=2) {
    x = x + 2;
}</pre>
```

provides a bigger window for local analysis

• What are some compiler optimizations you know about?

let's do a few more

Function inlining

```
int add(int x, int y) {
  return x + y;
}
int foo(int x, int y, int z) {
  return add(x,y);
}
```

```
int foo(int x, int y, int z) {
  return x + y;
}
```

What does this save us? code size? speed? the ability to debug? local regions to optimize more?

What are some compiler optimizations you know about?

There are many more! This is an active area of research and development

For a rough metric:

git effort shows activities on different files and directories

clang C++/C parser: 3.5K commits

clang AST: 8.7K commits

LLVM transforms/optimizations: 30K commits

The LLVM project has been under active development since 2000, and Clang since about 2007–2008.

The transformation part of the code base has the most activity by far

• How do you enable compiler optimizations?

- How do you enable compiler optimizations?
 - most C/++ compilers
 - optimizing for speed
 - -00, -01, -02, -03
 optimizing for size
 - -Os, -Oz
 - relax some constraints (especially around floating point):
 - Ofast
 - We can use Godbolt to test various optimizations.

STABILIZER: Statistically Sound Performance Evaluation

Charlie Curtsinger Emery D. Berger

Department of Computer Science University of Massachusetts Amherst Amherst, MA 01003 {charlie,emery}@cs.umass.edu

2013 research paper

"the performance impact of -03 over -02 optimizations is indistinguishable from random noise."

• What are some of the biggest improvements you've seen from compiler optimizations?

• What are some of the biggest improvements you've seen from compiler optimizations?

 compiler optimizations are great at well-structured, regular loops and arrays

• What kind of transforms on your code is the compiler allowed to do?

- What kind of transforms on your code is the compiler allowed to do?
- Matrix-add
- Many-add example

• Why did we get such a dramatic increase?

• Extreme example

```
void foo(int * arr, int n)
{
    int i, j;
    for (i = 0; i < n - 1; i++)
        for (j = 0; j < n - i - 1; j++)
        if (arr[j] > arr[j + 1]) {
            tmp = arr[j];
            arr[j] = arr[j + 1]);
            arr[j + 1] = tmp;
        }
}
```

```
int p(int arr[], int start, int end)
    int pivot = arr[start];
    int count = 0;
    for (int i = start + 1; i \leftarrow end; i++) {
        if (arr[i] <= pivot)</pre>
            count++;
    int pivotIndex = start + count;
    swap(arr[pivotIndex], arr[start]);
    int i = start, j = end;
    while (i < pivotIndex && j > pivotIndex) {
        while (arr[i] <= pivot) {</pre>
            i++:
        while (arr[j] > pivot) {
            j--;
        if (i < pivotIndex && j > pivotIndex) {
            swap(arr[i++], arr[j--]);
    return pivotIndex;
void foo(int *arr, int n)
    if (start >= end)
        return;
    int p = p(arr, m, n);
    foo(arr, start, p - 1);
    foo(arr, p + 1, end);
```

• Extreme example

```
bubble sort
```

```
void foo(int * arr, int n)
{
    int i, j;
    for (i = 0; i < n - 1; i++)
        for (j = 0; j < n - i - 1; j++)
        if (arr[j] > arr[j + 1]) {
            tmp = arr[j];
            arr[j] = arr[j + 1]);
            arr[j + 1] = tmp;
        }
}
```

```
Yes this transform would be legal!
```

Could any compiler figure it out? currently unlikely..

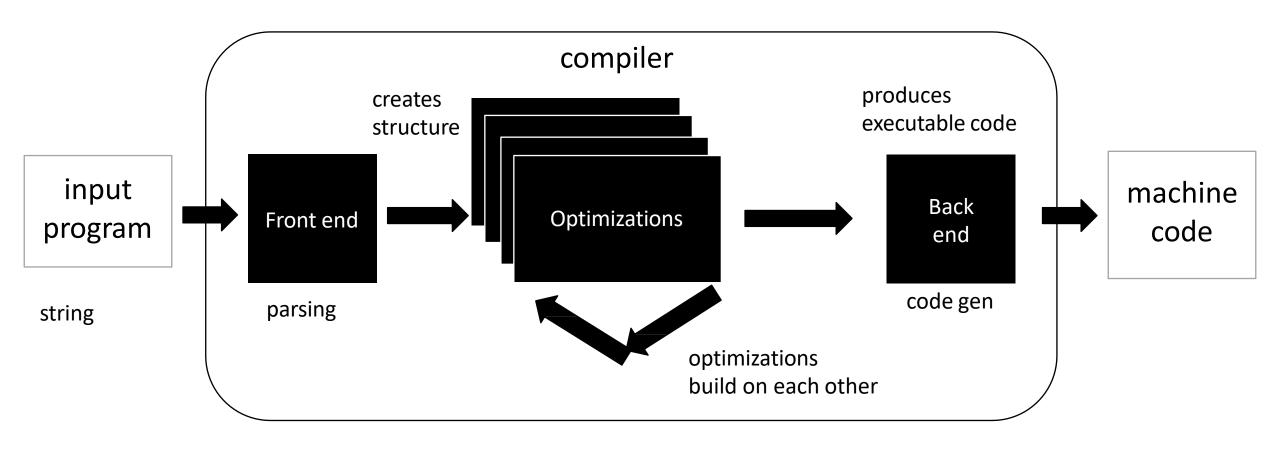
This is a technique called "super optimizing" and it is getting more and more interest

```
int p(int arr[], int start, int end)
    int pivot = arr[start];
    int count = 0;
    for (int i = start + 1; i \leftarrow end; i++) {
        if (arr[i] <= pivot)</pre>
             count++;
    int pivotIndex = start + count;
    swap(arr[pivotIndex], arr[start]);
    int i = start, j = end;
    while (i < pivotIndex && j > pivotIndex) {
        while (arr[i] <= pivot) {</pre>
             i++:
        while (arr[j] > pivot) {
            j--;
        if (i < pivotIndex && j > pivotIndex) {
             swap(arr[i++], arr[j--]);
    return pivotIndex;
void foo(int *arr, int n)
    if (start >= end)
        return;
    int p = p(arr, m, n);
    foo(arr, start, p - 1);
    foo(arr, p + 1, end);
```

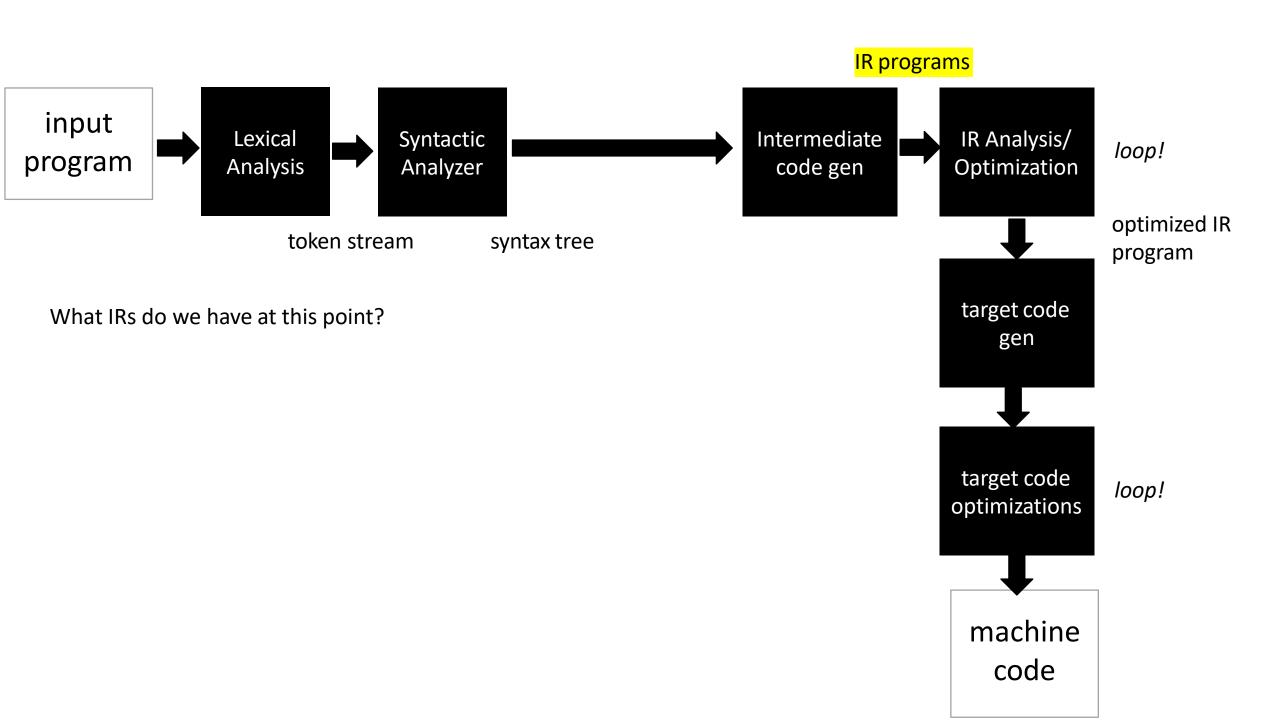
quick sort

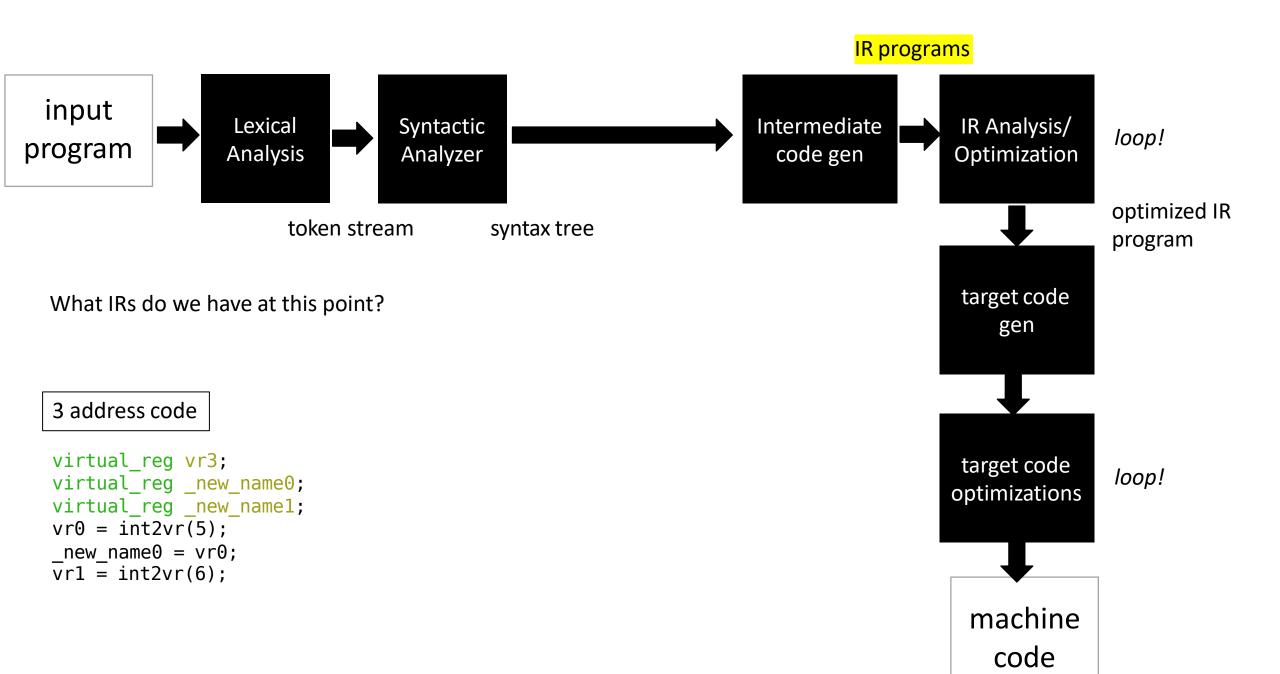
Moving on

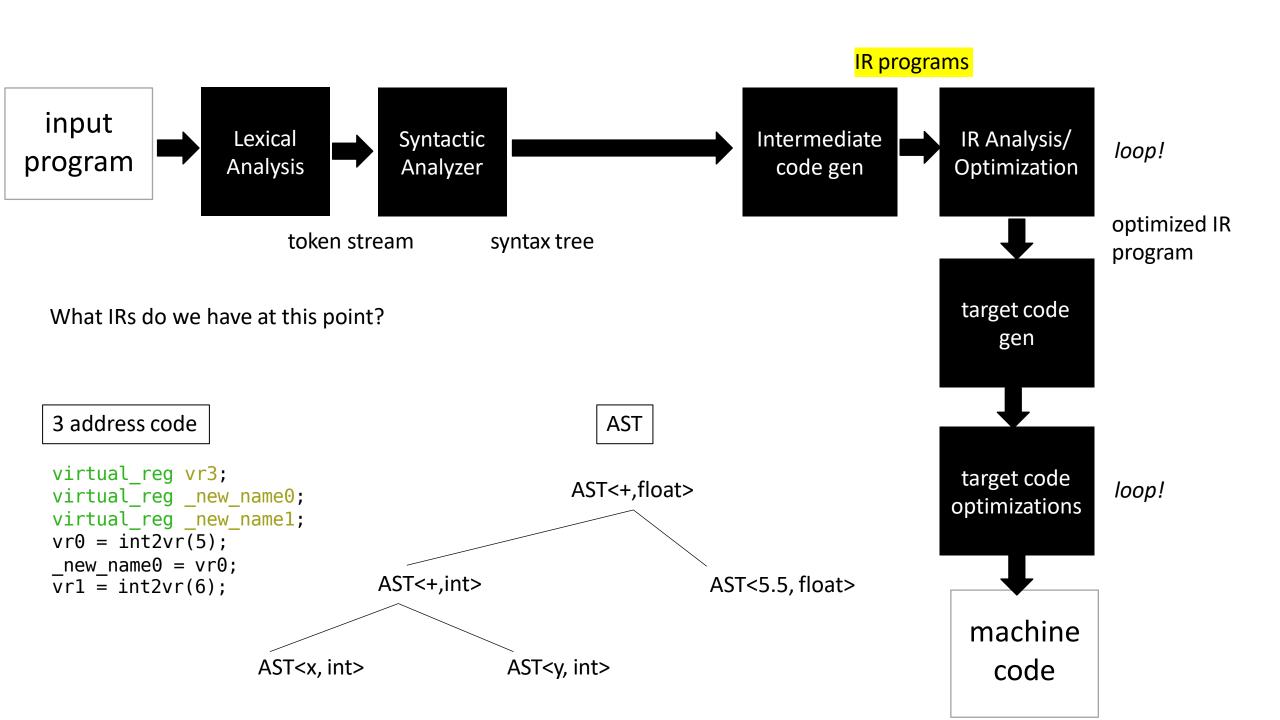
Zooming out again: Compiler Architecture



IRs and type inference type inference are at the boundary of parsing and optimizations





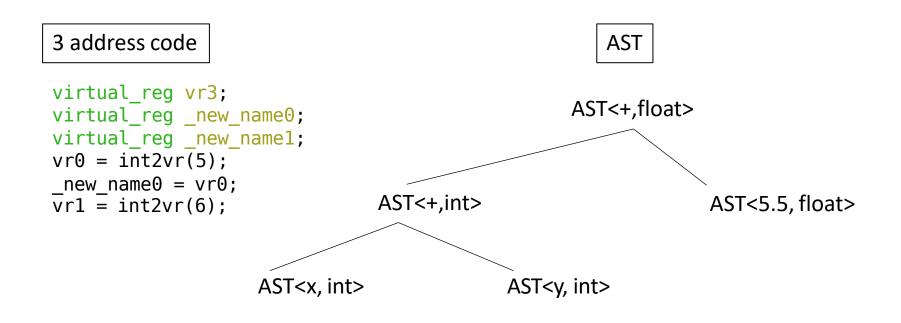


Implicit parse tree

```
if_else_statement := IF LPAR expr RPAR statement
if (program0) {
   program1
}
else {
   program2
}

We have several structures to utilize
to analyze and optimize programs!
```

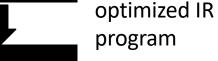
What IRs do we have at this point?



IR programs



loop!

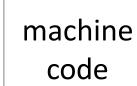




target code

optimizations

loop!



- Machine-independent these optimizations should work well across many different systems
 - Examples?

- Machine dependent these optimizations start to optimize the code for a given system
 - Examples?

- Machine-independent these optimizations should work well across many different systems
 - Examples?
 - All the examples we looked at before seem like they will help across many systems

- Machine dependent these optimizations start to optimize the code for a given system
 - Examples?
 - loop chunking for cache line size and vectorization.
 - instruction re-orderings to take advantage of processor pipelines.
 - fused multiply-and-add instructions

- Machine-independent these optimizations should work well across many different systems
 - Examples?
 - All the examples we looked at before seem like they will help across many systems

• In this module we will be looking at machine-independent optimizations.

What are the pros of machine-independent optimizations?

Next category level is how much code we need to reason about for the optimization.

- Local optimizations: examine a "basic block", i.e. a small region of code with no control flow.
 - Examples?
- Regional optimizations: several basic blocks with simple control flow.
 - Examples?
- Global optimization: optimizes across an entire function

- **Local optimizations**: examine a "basic block", i.e. a small region of code with no control flow.
- Regional optimizations: several basic blocks with simple control flow
- Global optimization: optimizes across an entire function
- IDFA Inter-Procedural Data Flow Analysis: Optimizes across functions

Discussion:

- What are the pros and cons of each?
- Going across functions is less common, why?

- **Local optimizations**: examine a "basic block", i.e. a small region of code with no control flow.
- Regional optimizations: several basic blocks with simple control flow
- Global optimization: optimizes across an entire function

For this module:

- We will look at two optimizations in detail:
- A local optimization: Local value numbering
- A regional optimization: Loop unrolling
- We will implement both as homework
- We will discuss several other optimizations and analysis

Basic blocks

IR Program structure

A sequence of 3 address instructions

- Programs can be split into **Basic Blocks**:
 - A sequence of 3 address instructions such that: there is a single entry, single exit

 Important property: an instruction in a basic block can assume that all preceding instructions will execute

Single Basic Block

```
Label_x:
op1;
op2;
op3;
br label_z;
```

IR Program structure

A sequence of 3 address instructions

- Programs can be split into Basic Blocks:
 - A sequence of 3 address instructions such that: there is a single entry, single exit

 Important property: an instruction in a basic block can assume that all preceding instructions will execute

Two Basic Blocks

```
Label_x:
op1;
op2;
op3;
br label_z;
```

Single Basic Block

IR Program structure

A sequence of 3 address instructions

- Programs can be split into Basic Blocks:
 - A sequence of 3 address instructions such that: there is a single entry, single exit

• Important property: an instruction in a basic block can assume that all preceding instructions will execute

How might they appear in a high-level language? What are some examples?

Two Basic Blocks

Single Basic Block

IR Program structure

A sequence of 3 address instructions

- Programs can be split into Basic Blocks:
 - A sequence of 3 address instructions such that: there is a single entry, single exit

 Important property: an instruction in a basic block can assume that all preceding instructions will execute How might they appear in a high-level language?

How many basic blocks?

```
...
if (x) {
    ...
}
else {
    ...
}
...
```

Two Basic Blocks

Single Basic Block

```
Label_x:
op1;
op2;
op3;
br label_z;
```

Converting 3 address code into basic blocks

• Let's try an example: test 4 in HW 4:

Converting 3 address code into basic blocks

- Simple algorithm:
 - keep a list of basic blocks
 - a basic block is a list of instructions
 - Iterate over the 3 address instructions
 - if you see a branch or a label, finalize the current basic block and start a new one.

Converting 3 address code into basic blocks

pseudo code

```
basic_blocks = []
bb = []
for instr in program:
    if instr type is in [branch, label]:
        bb.append(instr)
        basic_blocks.append[bb]
        bb = []
    else:
        bb.append(instr)
```

Local optimizations:

Optimizes an individual basic block

Regional optimizations:

Combines several basic blocks

- operates across an entire procedure
- what about across procedures?

Local optimizations:

Optimizes an individual basic block

Regional optimizations:

Combines several basic blocks

- operates across an entire procedure
- what about across procedures?

Local optimizations:

Optimizes an individual basic block

Regional optimizations:

Combines several basic blocks

- operates across an entire procedure
- what about across procedures?

Local optimizations:

Optimizes an individual basic block

Regional optimizations:

Combines several basic blocks

- operates across an entire procedure
- what about across procedures?

Label_0:

$$x = a + b$$
;

CANNOT
always optimized
to

Label_0:
 $x = a + b$;

Label_1:
 $y = a + b$;

 $y = x$;

Local optimizations:

Optimizes an individual basic block

Regional optimizations:

Combines several basic blocks

Global optimizations:

- operates across an entire procedure
- what about across procedures?

Label_0:

$$x = a + b;$$

 $y = a + b;$

optimized
to

Label_0:
 $x = a + b;$
 $y = x;$

Label_0:

$$x = a + b$$
;

CANNOT
always optimized
to

Label_0:
 $x = a + b$;

Label_1:
 $y = a + b$;

 $y = x$;

code could skip Label_0,
leaving x undefined!

Regional Optimization

```
...
if (x) {
    ...
}
else {
    x = a + b;
}
y = a + b;
...
```

```
we cannot replace:
    y = a + b.
    with
    y = x;
```

Regional Optimization: Data Flow Analysis

```
if (x) {
    if (x) {
        ...
}
else {
        x = a + b;
}
y = a + b;
...
```

```
we cannot replace:
    y = a + b.
    with
    y = x;
```

```
x = a + b;
if (x) {
    ...
}
else {
    ...
}
y = a + b;
```

Can x = a + b be hoisted out of the if then else?

```
In that case, we can check if a
and b are not redefined, then
y = a + b;
can be replaced with
y = x;
```

This requires regional analysis and optimizations

• A local optimization over 3 address code

 Attempts to replace arithmetic operations (expensive) with copy instructions (cheap)

• A local optimization over 3 address code

 Attempts to replace arithmetic operations (expensive) with copy instructions (cheap)

Can be extended to a regional optimization using flow analysis

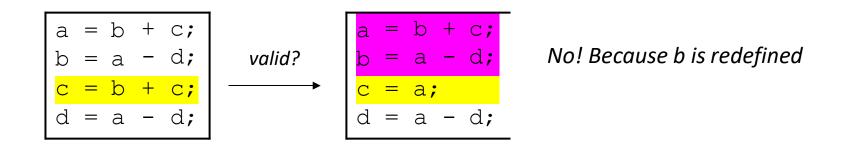
We are going to use notation similar to Class IR applicable to any Intermediate reapresentation.

• A local optimization over 3 address code

 Attempts to replace arithmetic operations (expensive) with copy instructions (cheap)

• A local optimization over 3 address code

 Attempts to replace arithmetic operations (expensive) with copy instructions (cheap)



• A local optimization over 3 address code

 Attempts to replace arithmetic operations (expensive) with copy instructions (cheap)

• A local optimization over 3 address code

 Attempts to replace arithmetic operations (expensive) with copy instructions (cheap)

$$a = b + c;$$

 $b = a - d;$
 $c = b + c;$
 $d = a - d;$
 $d = a - d;$
 $a = b + c;$
 $b = a - d;$
 $c = b + c;$
 $d = b + c;$
 $c = b + c;$
 $d = b + c;$
 $d = b + c;$

Local value numbering (LVN)

LVN Algorithm:

- Provide a number to each variable. Update the number each time the variable is updated.
- Keep a global counter; increment with new variables or assignments

Global counter = 0

Algorithm:

- Provide a number to each variable. Update the number each time the variable is updated.
- Keep a global counter; increment with new variables or assignments

$$a = b0 + c1;$$
 $b = a - d;$
 $c = b + c;$
 $d = a - d;$

Global_counter = 2

b and c have not been seen, give them a number

Algorithm:

- Provide a number to each variable. Update the number each time the variable is updated.
- Keep a global counter; increment with new variables or assignments

$$a2 = b0 + c1;$$
 $b = a2 - d3;$
 $c = b + c;$
 $d = a - d;$

Global_counter = 4

a2 have been updated use the last defined number. i.e. a on RHS, becomes a2 and so on.

Algorithm:

- Provide a number to each variable. Update the number each time the variable is updated (assigned to).
- Keep a global counter; increment with new variables or assignments

$$a2 = b0 + c1;$$
 $b4 = a2 - d3;$
 $c5 = b4 + c1;$
 $d6 = a2 - d3;$

Global_counter = 7

So ... for LHS variable always increment with updated number, for right hand side use, always appended last number version

Algorithm: Now that variables are numbered

 Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.

• At each step, check to see if the rhs has already been computed.

$$a2 = b0 + c1;$$
 $b4 = a2 - d3;$
 $c5 = b4 + c1;$
 $d6 = a2 - d3;$

If we remember to declare every one of these variable names, the IR is still valid!!!

Algorithm: Now that variables are numbered

• Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.

```
\begin{array}{c} \longrightarrow & a2 = b0 + c1; \\ b4 = a2 - d3; \\ c5 = b4 + c1; \\ d6 = a2 - d3; \end{array}
```

```
H = {
```

Algorithm: Now that variables are numbered

• Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.

```
\begin{array}{c} \longrightarrow & a2 = b0 + c1; \\ b4 = a2 - d3; \\ c5 = b4 + c1; \\ d6 = a2 - d3; \end{array}
```

```
H = {
      "b0 + c1" : "a2",
}
```

Algorithm: Now that variables are numbered

• Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.

```
\begin{array}{c} a2 = b0 + c1; \\ b4 = a2 - d3; \\ c5 = b4 + c1; \\ d6 = a2 - d3; \end{array}
```

```
H = {
      "b0 + c1" : "a2",
}
```

Algorithm: Now that variables are numbered

• Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.

```
a2 = b0 + c1;
b4 = a2 - d3;
c5 = b4 + c1;
d6 = a2 - d3;
```

```
H = \{
    "b0 + c1" : "a2",
    "a2 - d3" : "b4",
}
```

Algorithm: Now that variables are numbered

• Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.

```
a2 = b0 + c1;
b4 = a2 - d3;
c5 = b4 + c1;
d6 = a2 - d3;
```

Algorithm: Now that variables are numbered

• Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.

• At each step, check to see if the rhs has already been computed.

```
a2 = b0 + c1;
b4 = a2 - d3;
c5 = b4 + c1;
d6 = a2 - d3;
```

```
H = {
    "b0 + c1" : "a2",
    "a2 - d3" : "b4",
}
```

mismatch due to re-numbering!!!

i.e. it is no longer just c = b + c

Algorithm: Now that variables are numbered

• Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.

• At each step, check to see if the rhs has already been computed.

```
a2 = b0 + c1;
b4 = a2 - d3;
c5 = b4 + c1;
d6 = a2 - d3;
```

```
H = {
        "b0 + c1" : "a2",
        "a2 - d3" : "b4",
        "b4 + c1" : "c5",
}
```

Add new entry in Hash table with new Numbering.

Algorithm: Now that variables are numbered

• Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.

```
a2 = b0 + c1;
b4 = a2 - d3;
c5 = b4 + c1;
d6 = a2 - d3;
```

```
H = \{

"b0 + c1" : "a2",

"a2 - d3" : "b4",

"b4 + c1" : "c5",

Do a look up and ...
}
```

Algorithm: Now that variables are numbered

• Iterate sequentially through instructions. Keep a hash table of the rhs (numbered variables and operation) mapped to their lhs.

```
a2 = b0 + c1;
b4 = a2 - d3;
c5 = b4 + c1;
d6 = b4;
```

What else can we do?

What else can we do?

Consider this snippet:

```
a2 = c1 - b0;
f4 = d3 * a2;
c5 = b0 - c1;
d6 = a2 * d3;
```

Commutative operations

What is the definition of commutative?

Commutative operations

What is the definition of commutative?

$$x OP y == y OP x$$

What operators are commutative? Which ones are not?

Adding commutativity to local value numbering

• For commutative operators (e.g. + *), the analysis should consider a deterministic order of operands.

You can use variable numbers or lexigraphical order

Algorithm optimization:

```
H = \{
```

Algorithm optimization:

 for commutative operations, re-order operands into a deterministic order

cannot re-order because - is not commutative

```
→ a2 = c1 - b0;

f4 = d3 * a2;

c5 = b0 - c1;

d6 = a2 * d3;
```

Algorithm optimization:

```
\begin{array}{c} a2 = c1 - b0; \\ f4 = d3 * a2; \\ c5 = b0 - c1; \\ d6 = a2 * d3; \end{array}
```

Algorithm optimization:

 for commutative operations, re-order operands into a deterministic order

re-ordered because a2 < d3 lexigraphically

```
a2 = c1 - b0;

f4 = d3 * a2;

c5 = b0 - c1;

d6 = a2 * d3;
```

```
H = {
        "c1 - b0" : "a2",
        "a2 * d3" : "f4",
}
```

Algorithm optimization:

```
a2 = c1 - b0;
f4 = d3 * a2;
c5 = b0 - c1;
d6 = a2 * d3;
```

```
H = {
        "c1 - b0" : "a2",
        "a2 * d3" : "f4",
}
```

Algorithm optimization:

```
a2 = c1 - b0;
f4 = d3 * a2;
c5 = b0 - c1;
d6 = a2 * d3;
```

```
H = {
        "c1 - b0" : "a2",
        "a2 * d3" : "f4",
        "b0 - c1" : "c5",
}
```

Algorithm optimization:

```
a2 = c1 - b0;
f4 = d3 * a2;
c5 = b0 - c1;
d6 = a2 * d3;
```

```
H = {
        "c1 - b0" : "a2",
        "a2 * d3" : "f4",
        "b0 - c1" : "c5",
}
```

Algorithm optimization:

 for commutative operations, re-order operands into a deterministic order

And achieves more optimization opportunities ...

```
a2 = c1 - b0;
f4 = d3 * a2;
c5 = b0 - c1;
d6 = f4;
```

```
H = {
        "c1 - b0" : "a2",
        "a2 * d3" : "f4",
        "b0 - c1" : "c5",
}
```

Optimizing over wider regions

Local value numbering operated over just one basic block.

 We want optimizations that operate over several basic blocks (a region), or across an entire procedure (global)

- For this, we need Control Flow Graphs and Flow Analysis
 - We may have time to discuss this later in the module

Next:

• More optimizations!