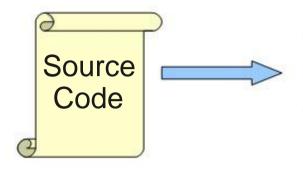
Three-Address Code IR

Where We Are



Lexical Analysis

Syntax Analysis

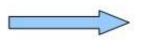
Semantic Analysis

IR Generation

IR Optimization

Code Generation

Optimization



Machine Code

Three-Address Code (TAC)

- High-level assembly language where each operation has at most three operands.
- Uses explicit runtime stack for function calls.
- When generating IR at this level, you do not need to worry about optimizing it.
- It's okay to generate IR that has lots of unnecessary assignments, redundant computations, etc.

Sample TAC Code

 Evaluating an expression sometimes requires the introduction of temporary variables.

Simple TAC Instructions

- Variable assignment allows assignments of the form
 - var = constant;
 - $var_1 = var_2;$
 - $var_1 = var_2 op var_3;$
 - $var_1 = constant op var_2;$
 - $var_1 = var_2 op constant;$
 - var = constant₁ op constant₂;
- Permitted operators are +, -, *, /, %
- How would you compile y = -x; ?

$$y = 0 - x;$$

One More with bools

```
int x;
int y;
bool b1;
bool b2;
bool b3;
b1 = x + x < y
b2 = x + x == y
b3 = x + x > y
```

'>' operator is not allowed.

TAC with bools

- Boolean variables are represented as integers that have zero or nonzero values.
- In addition to the arithmetic operator, TAC supports logical operators <, ==, | |, and &&.
- How might you compile $b = (x \le y)$?

Control Flow Statements

```
if (x < y)
   z = x;
else
   z = y;
z = z * z;</pre>
```

```
If x < y then t0 = 1
 Else t0 = 0
   IfZ t0 Goto L0;
   z = x;
   Goto L1;
LO:
   z = y;
```

Labels

- TAC allows for named labels indicating particular points in the code that can be jumped to.
- There are two control flow instructions:
 - Goto label;
 - IfZ value Goto label;
- Note that Ifz is always paired with Goto.

A Sample Program

```
void main() {
   int x, y;
   int m2 = x * x + y * y;

while (m2 > 5) {
      m2 = m2 - x;
   }
}
```

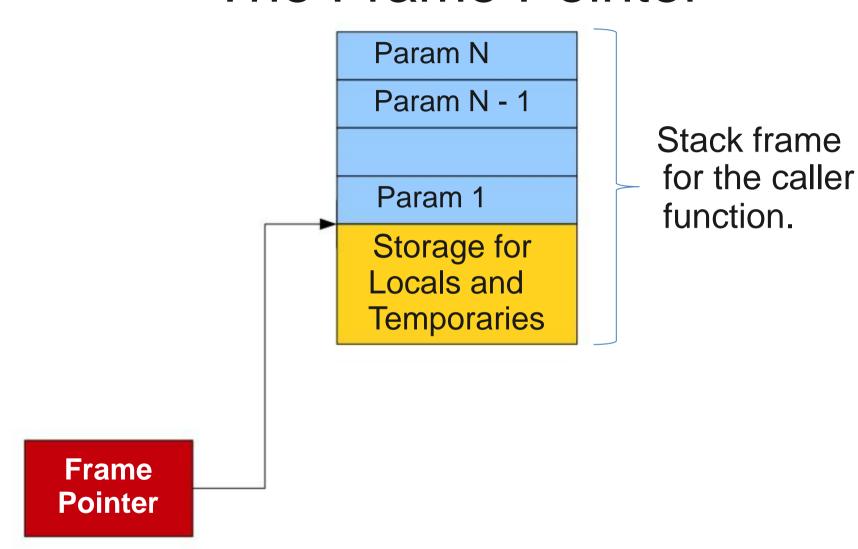
24 bytes for local variables and temporaries (6 * 4-bytes = 24 bytes) ► main: BeginFunc 24 t0 = x * x;t1 = y * y;m2 = t0 + t1;**LO**: t2 = 5 < m2;IfZ t2 Goto L1; m2 = m2 - x;Goto L0; EndFunc;

Compiling Function Calls

```
callee
void SimpleFn(int z) {
   int x, y;
   x = x * y * z;
}

void main() {
   Caller
   SimpleFn(137);
}
```

```
SimpleFn:
   BeginFunc 16;
    t0 = x * y;
   t1 = t0 * z;
   x = t1;
   EndFunc;
              Parameter z is
             already allocated
              by the caller
main:
   BeginFunc 4;
   t0 = 137;
   PushParam t0;
   LCall SimpleFn;
   PopParams 4;
   EndFunc;
```



Param N Param N - 1 Stack frame for the caller Param 1 function. Storage for Locals and **Temporaries** Param M Stack frame Param 1 **Frame** for the callee **Pointer** Storage for function. Locals and **Temporaries**

Param N

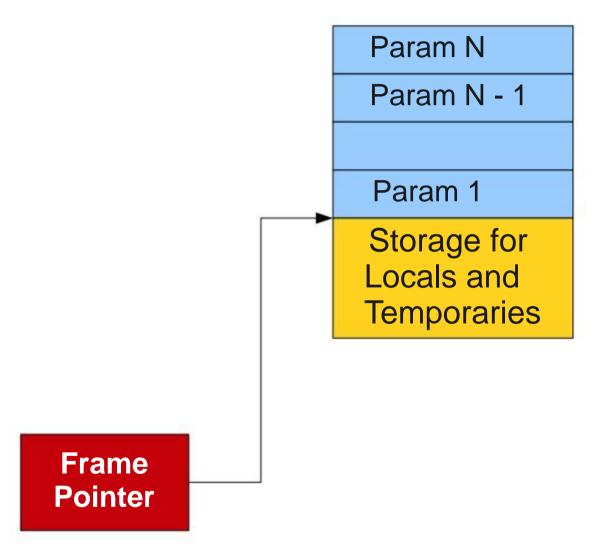
Param N - 1

Param 1

Storage for Locals and Temporaries

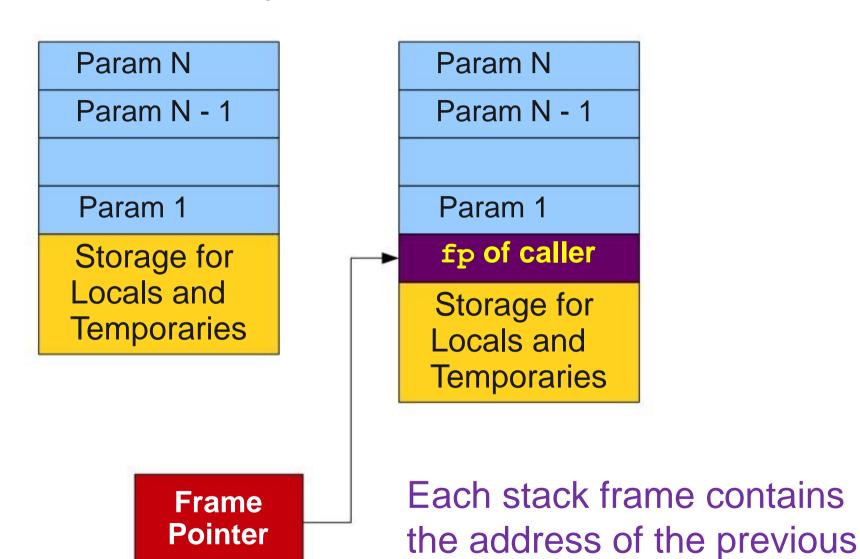
Stack frame for the caller function.



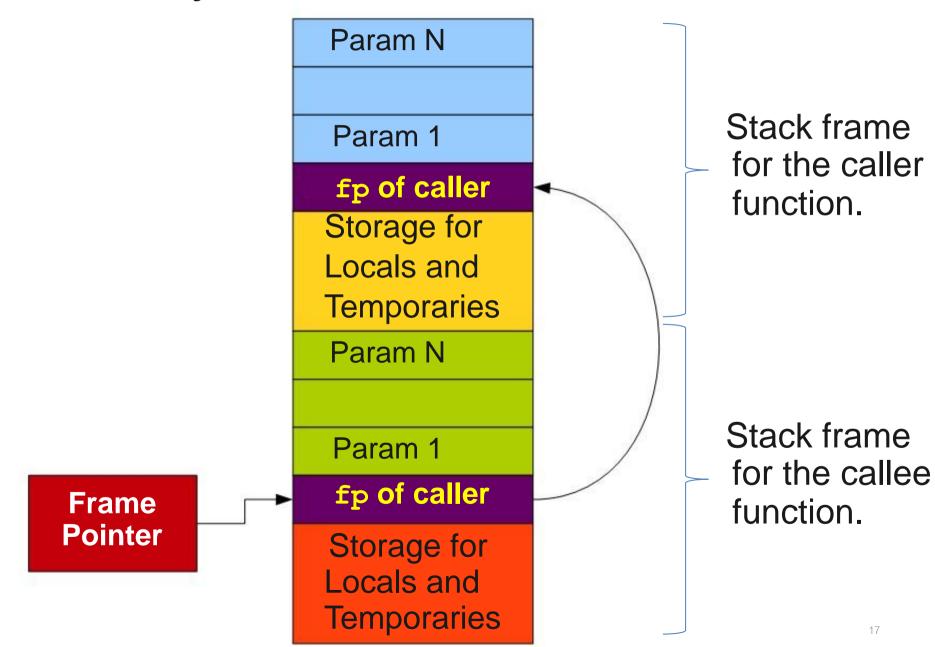


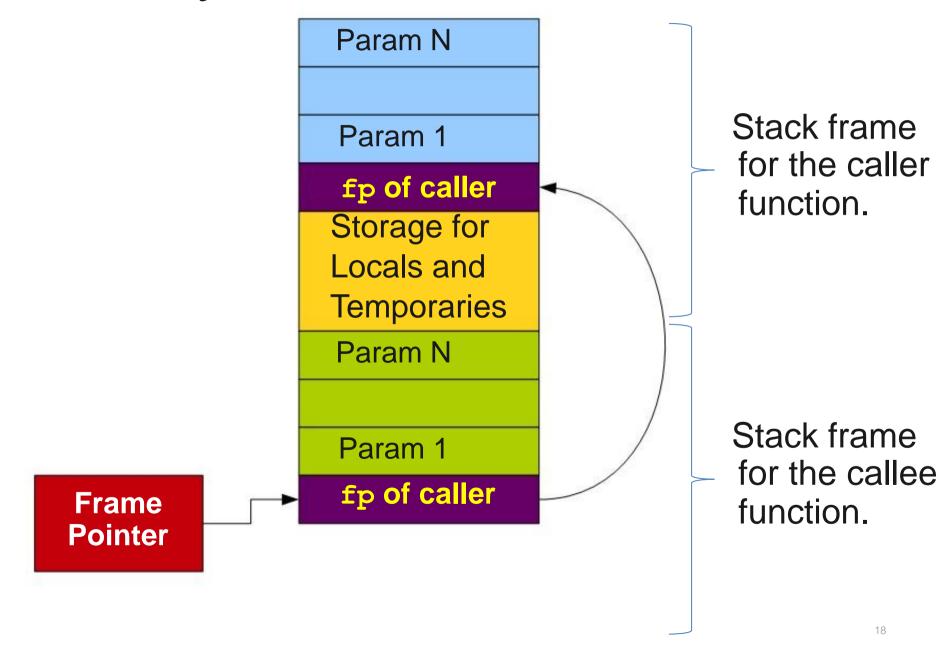
How do we know the position to move fp back?

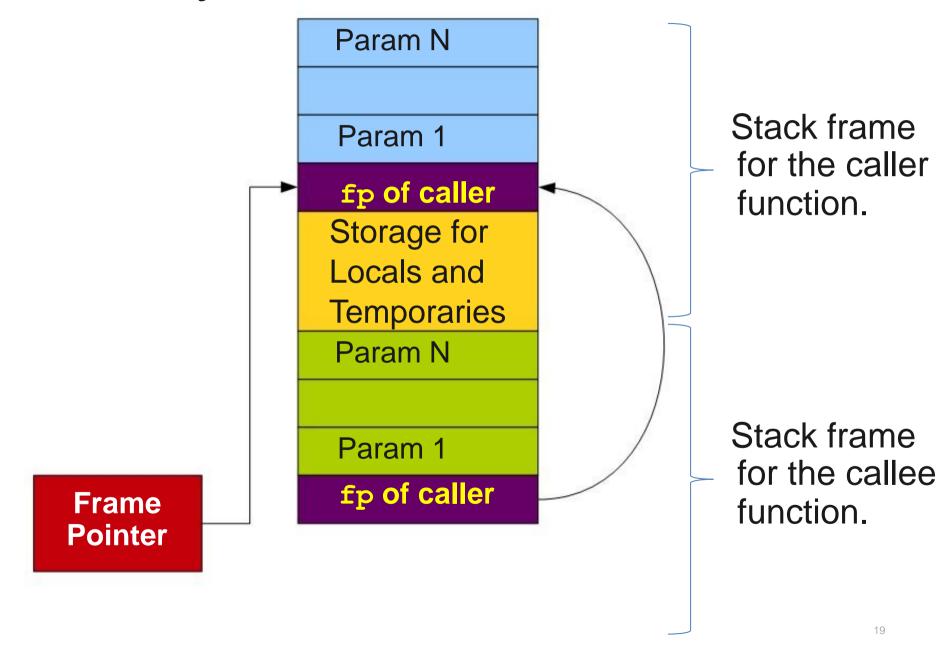
Logical vs Physical Stack Frames

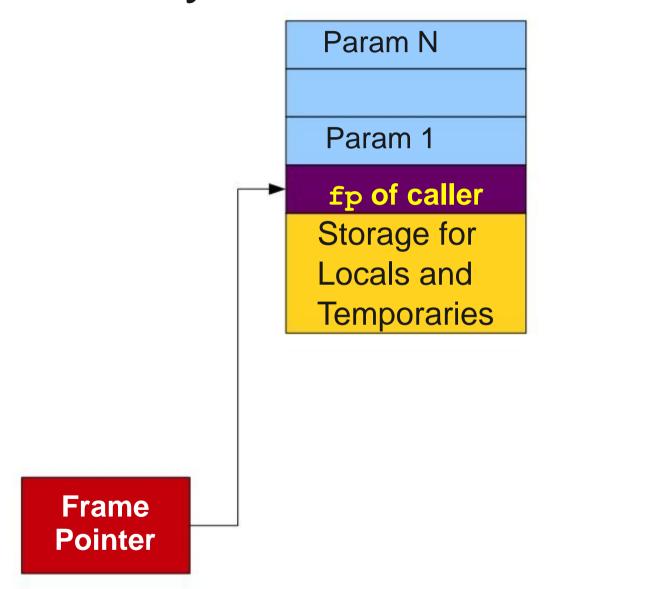


frame pointer.







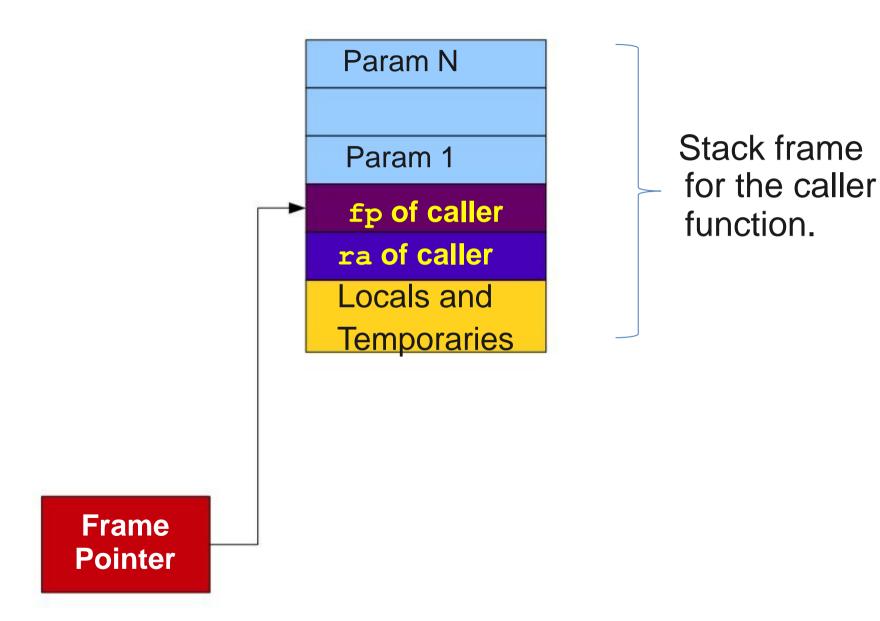


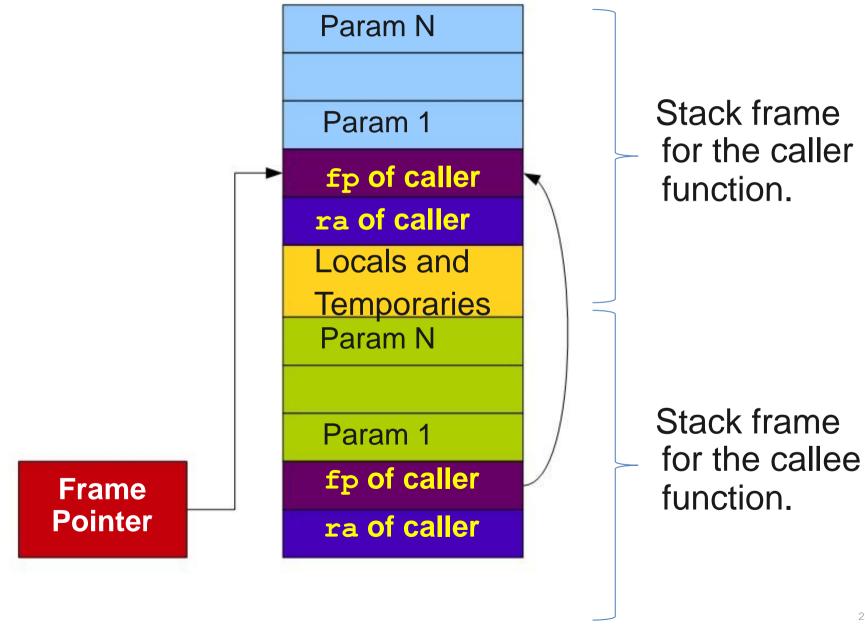
Stack frame for the caller function.

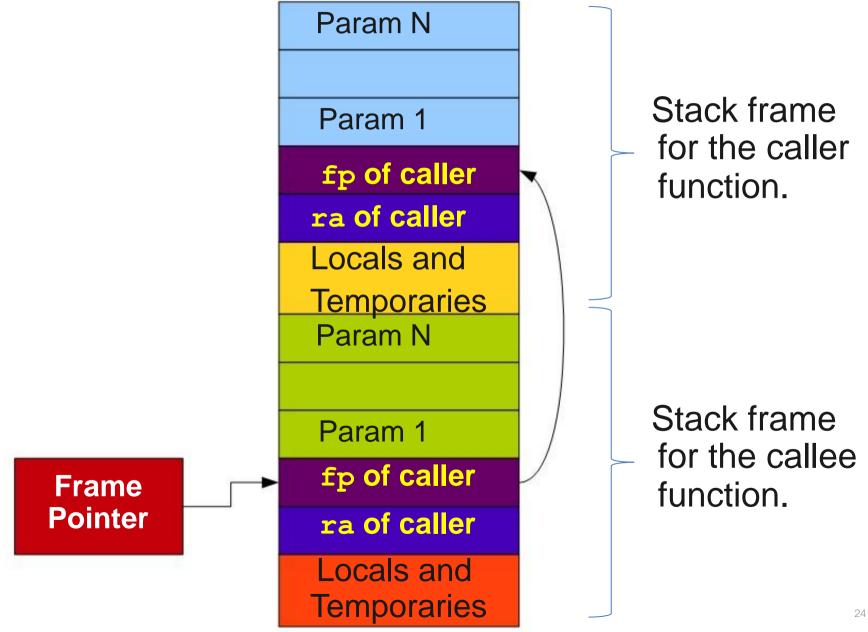
The Stored Return Address

- Internally, the processor has a special register called the program counter (PC) that stores the address of the next instruction to execute.
- Whenever a function returns, it needs to restore the PC so that the calling function resumes execution where it left off.
 - E.g. The address of where to return is stored in **MIPS*** in a special register called **ra** ("return address.")

^{*} MIPS (Microprocessor without Interlocked Pipeline Stages) is a reduced instruction set computer (RISC) instruction set (ISA) developed by MIPS Computer Systems, Inc. It compatible with both 32-Bit and 64-Bit architectures. It is widely used in small computing appliances, e.g. router, game station (PS2).





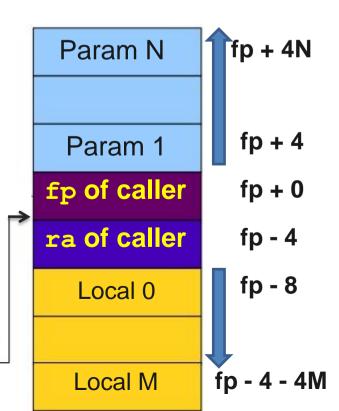


Locations of Parameters and Temp.

 Locations of local variable, parameter, and temporary variable are all relative to stack frame pointer (fp-relative).

Parameters begin at address
 fp + 4 and grow upward.

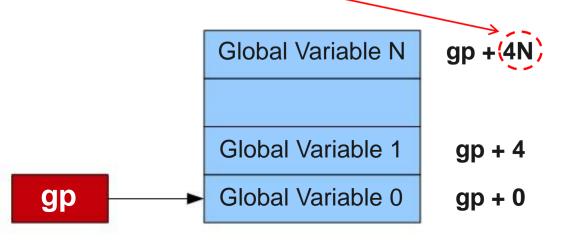
 Locals and temporaries begin at address fp - 8 and grow downward.



The Global Pointer

 The global pointer (gp) points to the memory treated as an array of values that grows upward.

 You must choose an offset into this array for each global variable.



Generating TAC

cgen for Atomic Expressions

```
cgen(c) = { // c is a constant
   Choose a new temporary t
   Emit( t = c);
   Return t
cgen(id) = { // id is an identifier}
   Choose a new temporary t
   Emit( t = id)
   Return t
```

cgen for Binary Operators

```
cgen(e_1 + e_2) = {
   Choose a new temporary t
   Let t_1 = cgen(e_1)
   Let t_2 = cgen(e_2)
   Emit( t = t_1 + t_2)
   Return t
```

```
cgen(5 + x) = {
   Choose a new temporary t
   Let t1 = cgen(5)
   Let t2 = cgen(x)
   Emit (t = t1 + t2)
   Return t
}
```

```
cgen(5 + x) = {
   Choose a new temporary t
   Let t_1 = \{
      Choose a new temporary t
      Emit(t = 5)
      return t
   Let t_2 = cgen(x)
   Emit (t = t_1 + t_2)
   Return t
```

```
cgen(5 + x) = {
   Choose a new temporary t
   Let t_1 = \{
      Choose a new temporary t
      Emit( t = 5 )
      return t
   Let t_2 = \{
      Choose a new temporary t
      Emit(t = x)
      return t
   Emit (t = t_1 + t_2)
   Return t
```

```
cgen(5 + x) = {
   Choose a new temporary t
   Let t_1 = \{
      Choose a new temporary t
      Emit( t = 5 )
      return t
                                      t3 = t1 + t2
   Let t_2 = \{
      Choose a new temporary t
      Emit(t = x)
      return t
   Emit (t = t_1 + t_2)
   Return t
```

cgen for Simple Statements

```
cgen(expr;) = {
   cgen(expr)
}
```

cgen for while loops

```
cgen(while (expr) stmt) = {
```

cgen for while loops

```
cgen(while (expr) stmt) = {
   Let L<sub>before</sub> be a new label.

Let L<sub>after</sub> be a new label.
```

```
cgen(while (expr) stmt) = {
   Let Lbefore be a new label.
   Let Lafter be a new label.
   Emit( Lbefore:)
   Emit( Lafter: )
```

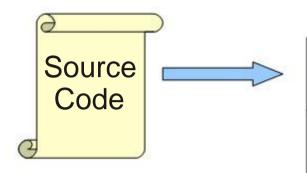
```
cgen(while (expr) stmt) = {
   Let L<sub>before</sub> be a new label.
   Let Lafter be a new label.
   Emit( Lbefore:)
   Let t = cgen(expr)
   Emit( IfZ t Goto Lafter )
   Emit( Lafter: )
```

```
cgen(while (expr) stmt) = {
   Let L<sub>before</sub> be a new label.
   Let Lafter be a new label.
   Emit( Lbefore:)
   Let t = cgen(expr)
   Emit(IfZ t Goto Lafter)
   cgen(stmt)
   Emit( Lafter: )
```

```
cgen(while (expr) stmt) = {
   Let L<sub>before</sub> be a new label.
   Let Lafter be a new label.
   Emit( Lbefore:)
   Let t = cgen(expr)
   Emit(IfZ t Goto Lafter)
   cgen(stmt)
   Emit(Goto Lbefore)
   Emit( Lafter: )
```

IR Optimization

Where We Are



Lexical Analysis

Syntax Analysis

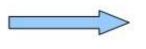
Semantic Analysis

IR Generation

IR Optimization

Code Generation

Optimization



Machine Code

Optimizations from IR Generation

```
int x;
int y;
bool b1;
bool b2;
bool b3;

b1 = x + x < y
b2 = x + x == y
b3 = x + x > y
```

```
t0 = x + x;
_{t1} = y;
b1 = t0 < t1;
 t2 = x + x;
t3 = y;
b2 = t2 == t3;
t4 = x + x;
t5 = y;
b3 = t5 < t4;
```

Optimizations from IR Generation

```
int x;
int y;
bool b1;
bool b2;
bool b3;

b1 = x + x < y
b2 = x + x == y
b3 = x + x > y
```

```
t0 = x + x;
b1 = t0 < t1;
 t2 = x + x;
b2 = t2 == _t3;
 t4 = x + x;
b3 = t5 < t4;
```

Optimizations from IR Generation

```
int x;
int y;
bool b1;
bool b2;
bool b3;

b1 = x + x < y
b2 = x + x == y
b3 = x + x > y
```

```
_t0 = x + x;
_t1 = y;
b1 = _t0 < _t1;
b2 = t0 == t1;
b3 = t0 < t1;
```

Optimizations from Lazy Coders

```
while (x < y + z) {
 x = x - y;
```

```
_L0:
_t0 = y + z;
_t1 = x < _t0;
IfZ _t1 Goto _L1;
x = x - y;
Goto _L0;
_L1:
```

Optimizations from Lazy Coders

```
while (x < y + z) {
 x = x - y;
}
```

This code never changes.

```
_t0 = y + z;
_t1 = x < _t0;
IfZ _t1 Goto _L1;
x = x - y;
Goto _L0;
_L1:
```

Optimizations from Lazy Coders

```
while (x < y + z) {
 x = x - y;
}
```

```
_t0 = y + z;
_L0:
_t1 = x < _t0;
IfZ _t1 Goto _L1;
x = x - y;
Goto _L0;
_L1:
```

The Challenge of Optimization

A good optimizer

- Should never change the observable behavior of a program.
- Should produce IR that is as efficient as possible.
- Should not take too long to process inputs.

What are we Optimizing?

What are some quantities we might want to optimize?

- Runtime (make the program as fast as possible at the expense of time and power)
- Memory usage (generate the smallest possible executable at the expense of time and power)
- Power consumption (choose simple instructions at the expense of speed and memory usage)
- Plus a lot more (minimize function calls, reduce use of floating-point hardware, etc.)

Semantics-Preserving Optimizations

 An optimization is semantics-preserving if it does not alter the semantics of the original program.

• Examples:

- Eliminating unnecessary temporary variables.
- Computing values that are known statically at compile-time instead of runtime.
- Evaluating constant expressions outside of a loop instead of inside.

A Formalism for IR Optimization

- Every phase of the compiler uses some tools:
 - Scanning (Lexer) uses regular expressions.
 - Parsing uses CFGs.
 - Semantic analysis uses proof systems and symbol tables.
- In optimization, we need a Control-Flow Graph to capture the structure of a program in a way amenable to optimization.

Visualizing IR

```
main:
    BeginFunc 40;
    tmp0 = LCall ReadInteger;
    a = tmp0;
    tmp1 = LCall ReadInteger;
    b = tmp1;
_L0:
    _{\text{tmp2}} = 0;
    tmp3 = b == tmp2;
    tmp4 = 0;
    tmp5 = tmp3 == tmp4;
    IfZ tmp5 Goto L1;
    c = a;
    a = b;
    tmp6 = c % a;
    b = tmp6;
    Goto L0;
L1:
    PushParam a;
    LCall PrintInt;
    PopParams 4;
    EndFunc;
```

Visualizing IR

Control Flow Graph

start

```
main:
                                     tmp0 = LCall ReadInteger;
   BeginFunc 40;
                                    \overline{a} = tmp0;
    tmp0 = LCall ReadInteger;
                                     tmp1 = LCall ReadInteger;
    a = tmp0;
                                    b = tmp1;
    tmp1 = LCall ReadInteger;
   b = tmp1;
_L0:
    tmp2 = 0;
                                     tmp2 = 0 ;
    tmp3 = b == tmp2;
                                     tmp3 = b == tmp2;
    tmp4 = 0;
                                    tmp4 = 0 ;
    tmp5 = tmp3 == tmp4;
                                     tmp5 = tmp3 == tmp4;
                                    IfZ tmp5 Goto L1;
    IfZ tmp5 Goto L1;
    c = a;
    a = b;
    tmp6 = c % a;
                                 c = a ;
    b = tmp6;
                                 a = b;
                                                   PushParam a ;
    Goto L0;
                                 tmp6 = c % a ;
                                                   LCall PrintInt;
L1:
                                 b = tmp6;
                                                   PopParams 4 ;
   PushParam a;
                                 Goto LO;
    LCall PrintInt;
    PopParams 4;
    EndFunc;
                                                          end
```

Basic Blocks & Control-Flow Graph

- A basic block is a sequence of IR instructions where
 - There is exactly one spot where control enters the sequence, which must be at the start of the sequence.
 - There is exactly one spot where control leaves the sequence, which must be at the end of the sequence.
- A control-flow graph (CFG) is a graph of the basic blocks in a function.
 - The term CFG is overloaded from here on out, we'll mean "control-flow graph" and not "contextfree grammar."

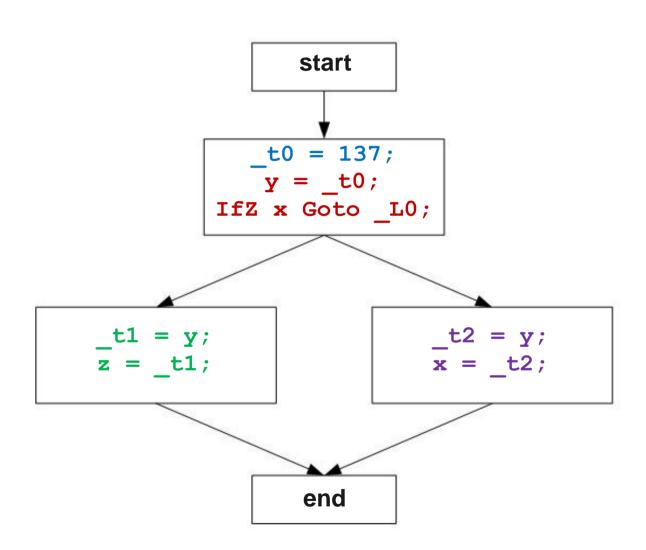
Types of Optimizations

- An optimization is local if it works on just a single basic block.
- An optimization is global if it works on an entire control-flow graph.
- An optimization is interprocedural if it works across the control-flow graphs of multiple functions. (We won't talk about this in this course.)

Example

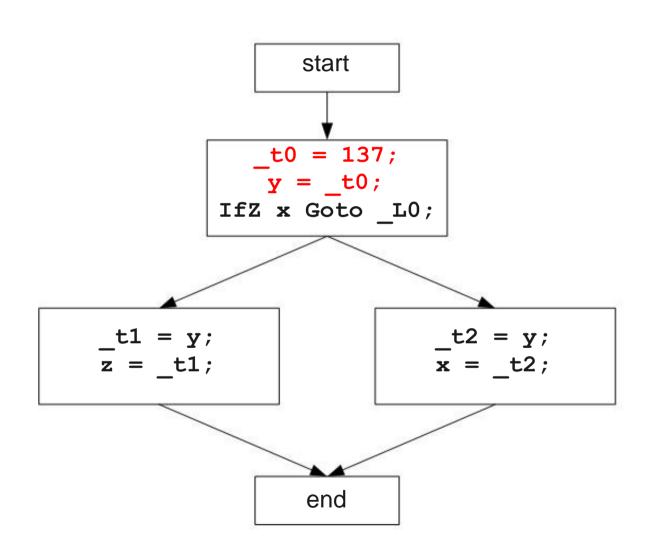
```
int main() {
    int x;
    int y;
    int z;

y = 137;
    if (x == 0)
    z = y;
    else
    x = y;
}
```



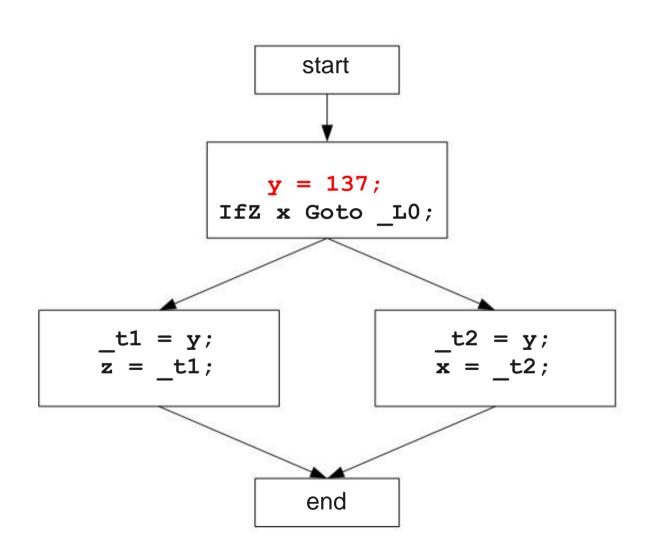
```
int main() {
    int x;
    int y;
    int z;

    y = 137;
    if (x == 0)
        z = y;
    else
        x = y;
}
```



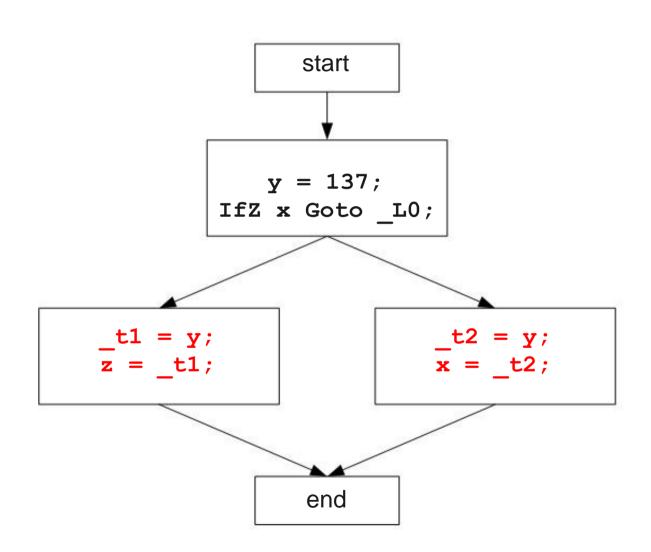
```
int main() {
    int x;
    int y;
    int z;

    y = 137;
    if (x == 0)
        z = y;
    else
        x = y;
}
```



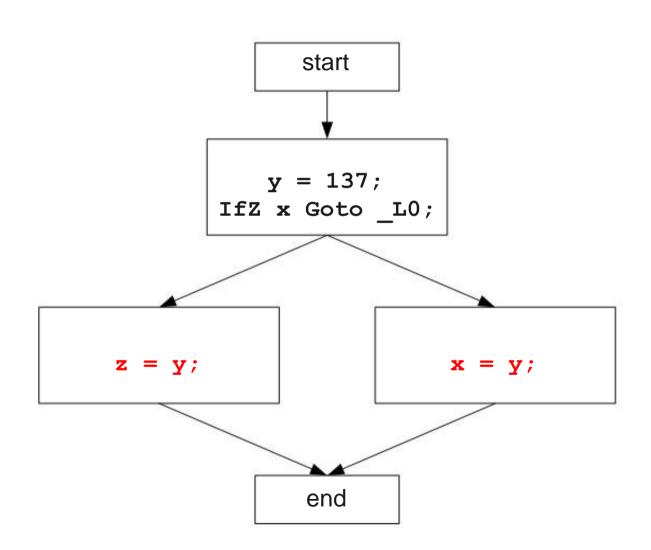
```
int main() {
    int x;
    int y;
    int z;

    y = 137;
    if (x == 0)
        z = y;
    else
        x = y;
}
```



```
int main() {
    int x;
    int y;
    int z;

    y = 137;
    if (x == 0)
        z = y;
    else
        x = y;
}
```

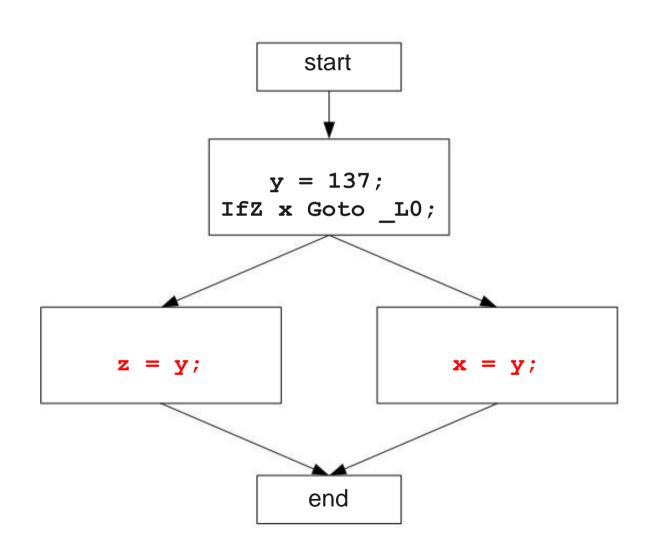


Global Optimizations

Example

```
int main() {
    int x;
    int y;
    int z;

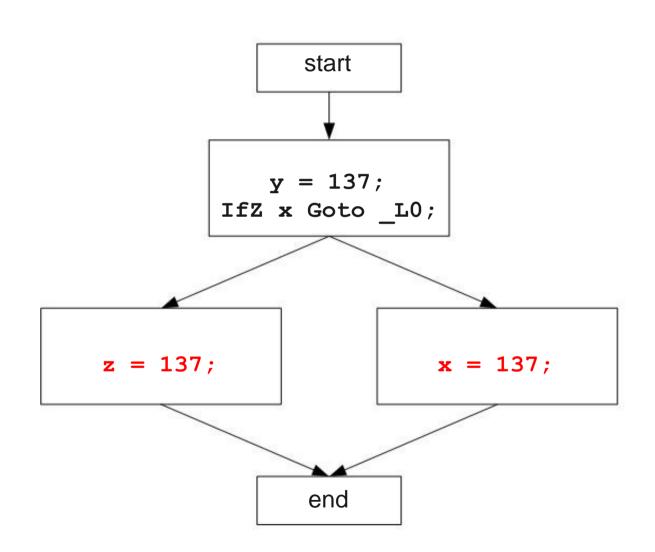
    y = 137;
    if (x == 0)
        z = y;
    else
        x = y;
}
```



Global Optimizations

```
int main() {
    int x;
    int y;
    int z;

    y = 137;
    if (x == 0)
        z = y;
    else
        x = y;
}
```



If we have two variable assignments

```
v_1 = a \text{ op } b
...

v_2 = a \text{ op } b
```

and the values of **v**₁, **a**, and **b** have not changed between the assignments, rewrite the code as

```
v_1 = a \text{ op } b
\dots
v_2 = v_1
```

Eliminates useless recalculation.

```
Object x;
int a;
int b;
int c;

x = new Object;
a = 4;
c = a + b;
x.fn(a + b);
```

```
tmp0 = 4 ;
PushParam tmp0;
tmp1 = LCall Alloc;
PopParams 4;
tmp2 = Object;
*(tmp1) = tmp2;
x = tmp1;
tmp3 = 4 ;
a = tmp3;
tmp4 = a + b ;
c = tmp4;
tmp5 = a + b ;
tmp6 = *(x) ;
tmp7 = *(tmp6);
PushParam tmp5;
PushParam x ;
ACall tmp7;
PopParams 8 ;
```

```
Object x;
int a;
int b;
int c;

x = new Object;
a = 4;
c = a + b;
x.fn(a + b);
```

```
tmp0 = 4 ;
PushParam tmp0 ;
tmp1 = LCall Alloc ;
PopParams 4 ;
tmp2 = Object;
\overline{*} ( tmp1) = tmp2;
x = tmp1;
tmp\bar{3} = 4 ;
a = tmp3;
tmp4 = a + b ;
c = tmp4;
tmp5 = a + b ;
tmp6 = *(x) ;
tmp7 = *(tmp6);
PushParam tmp5;
PushParam x ;
ACall tmp7;
PopParams 8 ;
```

```
Allocated
                      space
Object x;
              tmp1 stores
int a;
              allocated
             memory address
int b;
int c;
x = new Object;
a = 4;
c = a + b;
x.fn(a + b);
```

```
4 bytes
tmp0 = 4 \leftrightarrow
                   memory
                   address
PushParam tmp0 ;
tmp1 = LCall Alloc ;
PopParams 4 ;
tmp2 = Object;
*(tmp1) = tmp2;
x = tmp1;
tmp3 = 4 ;
a = tmp3;
tmp4 = a + b ;
c = tmp4;
 tmp5 = a + b ;
tmp6 = *(x) ;
tmp7 = *(tmp6);
PushParam tmp5;
PushParam x ;
ACall tmp7;
PopParams 8 ;
```

```
Object
tmp2
                 tmp1
       PTR to fn
                        Allocated
                         space
   Object x;
   int a;
   int b;
   int c;
   x = new Object;
   a = 4;
   c = a + b;
   x.fn(a + b);
```

```
tmp0 = 4 ;
PushParam tmp0;
tmp1 = LCall Alloc;
PopParams 4;
tmp2 = Object ;
*(tmp1) = tmp2;
x = tmp1;
tmp3 = 4 ;
a = tmp3;
tmp4 = a + b ;
c = tmp4;
tmp5 = a + b ;
tmp6 = *(x) ;
tmp7 = *(tmp6);
PushParam tmp5;
PushParam x ;
ACall tmp7;
PopParams 8 ;
```

```
Object
tmp2
                tmp1
      PTR to fn
  Object x;
  int a;
  int b;
   int c;
  x = new Object;
  a = 4;
  c = a + b;
  x.fn(a + b);
```

```
tmp0 = 4 ;
PushParam tmp0;
tmp1 = LCall Alloc;
PopParams 4;
tmp2 = Object;
*(tmp1) = tmp2;
x = tmp1;
tmp3 = 4 ;
a = tmp3;
tmp4 = a + b ;
c = tmp4;
tmp5 = a + b ;
tmp6 = *(x) ;
tmp7 = *(tmp6);
PushParam tmp5;
PushParam x ;
ACall tmp7;
PopParams 8 ;
```

```
Object
tmp2
                tmp1
      PTR to fn
  Object x;
   int a;
   int b;
   int c;
  x = new Object;
  a = 4;
  c = a + b;
  x.fn(a + b);
```

```
tmp0 = 4 ;
PushParam tmp0;
tmp1 = LCall Alloc;
PopParams 4;
tmp2 = Object;
*(tmp1) = tmp2;
x = tmp1;
tmp\overline{3} = 4 ;
a = tmp3;
tmp4 = a + b ;
c = tmp4;
tmp5 = a + b ;
tmp6 = *(x) ;
tmp7 = *(tmp6);
PushParam tmp5;
PushParam x ;
ACall tmp7;
PopParams 8 ;
```

```
Object
tmp2
      PTR to fn
  Object x;
  int a;
  int b;
  int c;
  x = new Object;
  a = 4;
  c = a + b;
  x.fn(a + b);
```

```
tmp0 = 4 ;
PushParam tmp0;
tmp1 = LCall Alloc;
PopParams 4;
tmp2 = Object;
*(tmp1) = tmp2;
x = tmp1;
tmp3 = 4;
a = tmp3;
 tmp4 = a + b;
c = tmp4;
 tmp5 = a + b ;
 tmp6 = *(x) ;
tmp7 = *(tmp6);
PushParam tmp5;
PushParam x ;
ACall tmp7;
PopParams 8 ;
```

```
Object
tmp2
       PTR to fn
   Object x;
   int a;
   int b;
   int c;
   x = new Object;
   a = 4;
   c = a + b;
   x.fn(a + b)
                     tmp7
         Object
           PTR to fn
                      Code for fn
```

```
tmp0 = 4 ;
PushParam tmp0;
tmp1 = LCall Alloc;
PopParams 4;
tmp2 = Object;
*(tmp1) = tmp2;
x = tmp1;
tmp3 = 4 ;
a = tmp3;
tmp4 = a + b ;
c = tmp4;
 tmp5 = a + b;
                  Follow
tmp6 = *(x) ;
                  pointer
 tmp7 = *(\_tmp6) \int ; to fn
PushParam tmp5 ;
PushParam x ;
ACall tmp7;
PopParams 8 ;
```

```
Object x;
int a;
int b;
int c;

x = new Object;
a = 4;
c = a + b;
x.fn(a + b);
```

```
tmp0 = 4 ;
PushParam tmp0;
tmp1 = LCall Alloc;
PopParams 4;
tmp2 = Object;
*(tmp1) = tmp2;
x = tmp1;
tmp3 = 4 ;
a = tmp3;
tmp4 = a + b ;
c = tmp4;
tmp5 = a + b ;
tmp6 = *(x) ;
tmp7 = *(tmp6);
PushParam tmp5;
PushParam x ;
ACall tmp7;
PopParams 8 ;
```

```
Object x;
int a;
int b;
int c;

x = new Object;
a = 4;
c = a + b;
x.fn(a + b);
```

```
tmp0 = 4 ;
PushParam tmp0;
tmp1 = LCall Alloc;
PopParams 4;
tmp2 = Object;
*(tmp1) = tmp2;
x = tmp1;
tmp3 = 4 ;
a = tmp3;
tmp4 = a + b ;
c = tmp4;
tmp5 = tmp4;
tmp6 = *(x) ;
tmp7 = *(tmp6);
PushParam tmp5;
PushParam x ;
ACall tmp7;
PopParams 8 ;
```

```
Object x;
int a;
int b;
int c;

x = new Object;
a = 4;
c = a + b;
x.fn(a + b);
```

```
tmp0 = 4 ;
PushParam tmp0;
tmp1 = LCall Alloc;
PopParams 4;
tmp2 = Object;
*(tmp1) = tmp2;
x = tmp1;
tmp3 = 4 ;
a = tmp3;
tmp4 = a + b ;
c = tmp4;
tmp5 = tmp4;
tmp6 = *(x);
tmp7 = *(tmp6);
PushParam tmp5;
PushParam x ;
ACall tmp7;
PopParams 8 ;
```

```
Object x;
int a;
int b;
int c;

x = new Object;
a = 4;
c = a + b;
x.fn(a + b);
```

Next class: More optimization on this code

```
tmp0 = 4 ;
PushParam tmp0;
tmp1 = LCall Alloc;
PopParams 4;
tmp2 = Object;
*(tmp1) = tmp2;
x = tmp1;
tmp3 = tmp0;
a = tmp3;
tmp4 = a + b ;
c = tmp4;
tmp5 = tmp4;
tmp6 = *(x);
tmp7 = *(tmp6);
PushParam tmp5;
PushParam x ;
ACall tmp7;
PopParams 8 ;
```